

## Federal Interagency Forum on Aging-Related Statistics

# PRE-PUBLICATION VERSION, DATA SOURCES ON OLDER AMERICANS 2008

**Pre-Publication Version, Data Sources on Older Americans 2008 (DSOA)** highlights the contents of government-sponsored surveys and products containing statistical information about the older population. All Federal agencies are invited to contribute to this report and participate in the Forum. More information about the Forum can be found at [www.agingstats.gov](http://www.agingstats.gov). The Forum plans to release an up-to-date product titled *Data Sources on Older Americans 2009* in the fall of 2009. **Pre-Publication Version, Data Sources on Older Americans 2008** is meant to serve as a temporary or intermediary product between the 2006 and 2009 versions. Not all data sources have been updated.

Questions about the **Pre-Publication Version, Data Sources on Older Americans 2008** may be directed to the Federal Interagency Forum on Aging-Related Statistics, National Center for Health Statistics, 3311 Toledo Road, Room 6321, Hyattsville, MD, 20782, phone: 301-458-4460, or e-mail: [agingforum@cdc.gov](mailto:agingforum@cdc.gov).

## Note to the User

The periodic summary **Pre-Publication Version, Data Sources on Older Americans 2008** highlights the aging-related products currently available from member agencies of the Federal Interagency Forum on Aging-Related Statistics (Forum) as well as other Federal agencies. Some data bases or surveys could be listed under more than one agency. Federal agencies often jointly develop data, but produce reports that reflect differing agency missions. For example, the Current Population Survey (CPS) is sponsored by the Bureau of Labor Statistics (BLS), and fielded by the U.S. Census Bureau. Reports based on CPS data and issued by the two agencies differ. BLS reports focus on employment and labor force topics, while U.S. Census Bureau reports focus on living arrangements, money, income, and poverty topics.

Questions about particular data sources or products should be directed to the agency contact specified. For additional copies of **Pre-Publication Version, Data Sources on Older Americans 2008**, contact the Federal Interagency Forum on Aging-Related Statistics, National Center for Health Statistics, 3311 Toledo Road, Room 6321, Hyattsville, MD, 20782, phone: 301-458-4460, or e-mail: [agingforum@cdc.gov](mailto:agingforum@cdc.gov).

## Table of Contents

Department of Commerce.....	1
U.S. Census Bureau .....	1
American Community Survey (ACS).....	1
Decennial Census (Census).....	4
International Data Base (IDB) .....	8
Longitudinal Employer-Household Dynamics (LEHD).....	10
National Longitudinal Mortality Study (NLMS).....	12
Population Projections .....	14
Survey of Income and Program Participation (SIPP) .....	15
Department of Health and Human Services.....	18
Administration on Aging .....	18
Census 2000 Special Tabulation on Aging.....	18
Agency for Healthcare Research and Quality.....	19
Healthcare Cost and Utilization Project (HCUP) .....	19
Medical Expenditure Panel Survey (MEPS).....	24
Centers for Disease Control and Prevention.....	28
National Center for Chronic Disease Prevention and Health Promotion .....	28
Behavioral Risk Factor Surveillance System (BRFSS) .....	28
National Center for Health Statistics .....	30
Longitudinal Studies of Aging (LSOAs).....	30
National Ambulatory Medical Care Survey (NAMCS).....	33
National Death Index (NDI) .....	35
National Health and Nutrition Examination Survey (NHANES).....	36
National Health and Nutrition Examination Survey—Epidemiologic Followup Study (NHEFS) .....	39
National Health Interview Survey (NHIS) .....	41
National Home and Hospice Care Survey (NHHCS).....	43
National Hospital Ambulatory Medical Care Survey (NHAMCS) .....	45
National Hospital Discharge Survey (NHDS) .....	47
National Mortality Followback Survey (NMFS).....	49
National Nursing Home Survey (NNHS) .....	51
National Nursing Home Survey Follow-up (NNHSF) .....	54
National Survey of Ambulatory Surgery (NSAS) .....	56
National Vital Statistics System (Vital Statistics) .....	57
Health Data Interactive (formerly Trends in Health and Aging).....	58
Centers for Medicare & Medicaid Services.....	59
Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surveys) .....	59
Medicaid (Claims and Enrollment) Data .....	60
Medicare (Fee-for-Service (FFS) Claims, Part D Events, and Parts A/B/C/D Enrollment) Data .....	62
Medicare Current Beneficiary Survey (MCBS) .....	65
Medicare Health Outcomes Survey (HOS).....	67
Minimum Data Set (MDS).....	70
Outcome and Assessment Information Set (OASIS).....	71

Section 723 Chronically Ill Medicare Beneficiary Research and Data Initiative (Section 723) .....	72
National Institutes of Health .....	74
National Cancer Institute .....	74
Surveillance, Epidemiology, and End Results (SEER) .....	74
National Institute on Aging.....	76
Health and Retirement Study (HRS).....	76
National Long Term Care Survey (NLTC) .....	80
New Immigrant Survey (NIS).....	83
Panel Study of Income Dynamics (PSID) .....	85
Wisconsin Longitudinal Study (WLS).....	88
Wisconsin National Survey of Midlife Development in the United States: MIDUS 1 (1995-1996) and MICUS 2 (2004-2005) .....	90
National Study of Assisted Living for the Frail Elderly, 2000.....	92
Substance Abuse and Mental Health Services Administration.....	93
Client/Patient Sample Survey (CPSS) .....	93
Drug Abuse Warning Network (DAWN).....	95
National Survey on Drug Use and Health (NSDUH) .....	98
National Survey of Substance Abuse Treatment Services (N-SSATS).....	101
National Survey of Mental Health Treatment Facilities (NSMHTF), formerly The Survey of Mental Health Organizations (SMHO).....	104
Treatment Episode Data Set (TEDS).....	106
Department of Housing and Urban Development .....	108
Office of Policy Development and Research.....	108
American Housing Survey (AHS) .....	108
Department of Labor.....	110
Bureau of Labor Statistics.....	110
American Time Use Survey (ATUS).....	110
Consumer Expenditure Survey (CEX).....	112
Current Population Survey (CPS).....	114
National Longitudinal Surveys (NLS) .....	116
Department of Veterans Affairs.....	119
National Survey of Veterans (NSV) .....	119
Survey of Veteran Enrollees' Health and Reliance Upon VA, 2007.....	120
Veteran Population Estimates and Projections (model name is VetPop2007, Dec. 2007) .....	121
Federal Reserve System, Board of Governors.....	122
Survey of Consumer Finances (SCF) .....	122
Office of Management and Budget.....	124
FedStats.....	124
Social Security Administration .....	126
New Beneficiary Data System (NBDS).....	126

List of Tables

Table 1. Summary of the National Longitudinal Survey Cohorts .....	116
--	-----

**Department of Commerce  
U.S. Census Bureau**

**American Community Survey (ACS)**

Sponsor: U.S. Census Bureau

Data Collection Agency/Organization: U.S. Census Bureau

**Purpose:** The American Community Survey (ACS) is a nationwide survey designed to provide communities with a fresh look at how they are changing. It will replace the long form in the 2010 Census. The ACS collects information from those in U.S. households and group quarters similar to what was collected on the Census 2000 long form, such as income, educational attainment, place of birth, disability status, home value, veteran status, and other important data. In addition to producing annual and multiyear characteristics of the population and housing, the survey also collects information for small areas such as blocks, groups, and population subgroups.

**Survey Universe:** A national sample of about 3 million addresses in every county, American Indian and Alaska Native area, and Hawaiian Home Land in the United States, and in Puerto Rico living in housing units and group quarters.

**Research Design:** Large continuous demographic survey.

**Survey Mode:** Paper and pencil questionnaire format. Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Personal Interviewing (CAPI) modes are used to follow-up with non-responding households.

**Unit of Analysis:** Person-level analysis.

**Sample:** In the 2006 ACS release, just over 7,000 geographic areas will receive data. These areas are large enough to meet the 65,000 population threshold for 1-year period estimates. In addition to the 1-year estimates, the Census Bureau plans to release 3-year period estimates, in 2007. These estimates will allow for the release of data for more geographic areas – areas with a population of 20,000 or more.

The Census Bureau conducts the ACS in every county of the United States—contacting the residents of 3 million housing units, as well as, the population living in group quarters. Group quarters (GQ) include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, and workers’ dormitories. The ACS universe is the entire resident population of the United States.

The 2006 ACS sample uses a two-stage stratified annual sample of approximately 838,000 housing units designed to measure socioeconomic and demographic characteristics of housing units and their occupants. 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 was selected from each stratum. The probability of selection for each PSU in the stratum is proportional to its estimated 2006 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 that are systematically sorted lists of addresses of housing units drawn from the MAF.

The 2006 sample for group quarters facilities included such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers' dormitories, and facilities for people experiencing homelessness. Each group quarters facilities is classified according to its group quarters type. All group quarters samples are selected during the main sampling phase. The sample is selected by state, not county, as is done with the housing unit sample. Some locations that were classified in Census 2000 as specified group quarters types are not included in ACS data collection operations. These are domestic violence shelters, soup kitchens, regularly scheduled mobile food vans, targeted non-sheltered outdoor locations, crews of commercial maritime vessels, natural disaster shelters, and dangerous encampments. There are several reasons for their exclusion, depending on the group quarters type, including concerns about privacy and the operational feasibility of repeating interviews for a continuing survey rather than once a decade for a census. Nevertheless, ACS estimates for the total population are controlled to be consistent with the intercensal population estimates. The exclusion of these group quarters types may result in a small bias in some ACS estimates.

Topics: Social characteristics, housing, income, National and small area estimates.

Data Availability: Since the American Community Survey reached full implementation in January 2005, data availability is on a program schedule. Similar to the Census website, The ACS website has various tools for accessing survey data and data products: <http://www.census.gov/acs/www/index.html>. Data tools include the fact sheets, data profiles, multiyear profiles, selected population profiles, and various maps.

Linked Data: N/A

Data Dissemination: Public-Use Microdata files from the ACS show the full range of responses made on individual questionnaires. All identifying information is removed to ensure confidentiality. The records selected are a sample of those households and group quarters that received the questionnaire. The questionnaire included questions on age, sex, tenure, income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, and other subjects. Also, the full range of population and housing information collected in the ACS is available on the Public-Use Microdata website for the ACS located at: <http://www.census.gov/acs/www/Products/PUMS/>.

PowerPoint presentations covering several topics on the ACS are also available through the Census Bureau's Public-Use Presentation Library, which is a collection of downloadable PowerPoint presentations, including slides and speaker notes, designed to inform data users about our various products and services. The web address for the Public-Use Presentation Library is: [http://www.census.gov/mso/www/pres\\_lib/index2.html#ACS](http://www.census.gov/mso/www/pres_lib/index2.html#ACS).

Reports: See: <http://www.census.gov/acs/www/>.

Future Plans: In 2008 the Census Bureau will release the first set of multiyear estimates for the nation. These 3-year estimates will be based on data collected from 2005 – 2007 and will include geographic areas with populations of 20,000 or more. In 2010 5-year estimates will be released based on data collected from 2005 – 2009 and will include the smallest geographic areas – down to the tract and block group levels.

For more information:

Email: [cmo.acs@census.gov](mailto:cmo.acs@census.gov)

Phone: 800-923-8282

Website: <http://www.census.gov/acs/www/index.html>

## **Decennial Census (Census)**

Sponsor: U.S. Census Bureau

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: The U.S. decennial census serves two main purposes: (1) to apportion the 435 seats in the U.S. House of Representatives among the 50 States—The U.S. Constitution, Article I, Section 2, apportionment of representatives among the States, for the House of Representatives, must be carried out every 10 years (decennially); and (2) to enumerate the resident population. For Census 2000, data on sex, race, Hispanic origin, age, and tenure were collected from 100 percent of the enumerated population. More detailed information, such as income, education, housing, occupation, and industry, was collected from a representative sample of the population.

Survey Universe: U.S. resident population.

Research Design: Census 2000 was the last count of the U.S. population collected by the Census Bureau. The Census Bureau's primary method of data collection is to mail out questionnaires using the Local Update of Census Addresses (LUCA) program and by using enumerators. Enumerators are Census Bureau staff that travel door-to-door gathering data by canvassing roads and streets looking for living quarters. For Census 2000, as in several previous censuses, two forms were used—a short form and a long form. The short form was sent to every household, and the long form, containing the seven 100 percent questions plus the sample questions, was sent to only a limited number of households—about one in every six homes. The extended census form collects information on social, housing, economic, and financial characteristics. The national final response rate for Census 2000 was 67 percent. This exceeded the projected response rate of 61 percent and was better than the 65 percent response rate from the 1990 census.

Survey Mode: One of two different survey forms was used to enumerate the U.S. population: (1) A short form with seven basic questions, (2) a long form including all questions from the short form and additional inquiry questions. On average, one in every six households received the long form.

Unit of Analysis: Person-level data analysis.

Sample: Every person and housing unit in the United States was asked basic demographic and housing questions (for example, race, age, and relationship to householder). A sample of these people and housing units was asked more detailed questions about items, such as income, occupation, and housing costs. The sampling unit for Census 2000 was the housing unit, including all occupants. There were found different housing unit sampling rates: 1-in-8, 1-in-6, 1-in-4 and 1-in-2 (designed for an overall average of about 1-in-6). The Census Bureau assigned these varying rates based on precensus occupied housing unit estimates of various geographic and statistics entities, such as incorporated places and interim census tracts. For people living in



group quarters (GQ) or enumerated at long form eligible sites (shelters and soup kitchens), the sampling unit was the person and the sampling rate was 1-in-6.

Topics: Population and housing characteristics, National, State, and local levels of geography.

Data Availability: In addition to conducting the Census every 10 years, updates to Census population counts are also conducted between census years.

Postcensal Population Estimates: These 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 component 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 U.S. civilian citizens

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 because 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.

Linked Data: See description of National Longitudinal Mortality Data (NLMS).

Data Dissemination: Data from Census 2000 and previous census surveys can be obtained primarily through various tools used on the Census website (<http://www.census.gov/main/www/cen2000.html>). Census 2000 is the first census for which the internet site listed above is the primary means of disseminating the data. In addition to formatted tables, the Census Bureau website has maps and data sets available for downloading (file transfer protocol (ftp)), printing, viewing, and manipulating. Special reports and briefs on Census data that provide background information, explain how data was analyzed, and differences between 1990 and 2000 data, can be obtained through the following website: <http://www.census.gov/population/www/cen2000/briefs/index.html>.

Public-Use Microdata Area (PUMA). A geographic entity for which the U.S. Census Bureau provides specially selected extracts of raw data from a small sample of long-form census records that are screened to protect confidentiality of census records. The extract files are referred to as public use microdata samples (PUMS). Public use microdata areas (PUMAs), which must have a minimum census population of 100,000 and cannot cross a State line, receive a 5-percent sample of the long form records; these records are presented in State files. These PUMAs are

aggregated into super-PUMAs, which must have a minimum census population of 400,000 and receive a 1-percent sample in a national file. PUMAs and super-PUMAs are mutually exclusive, that is, they use different records to create each sample. Data users can use these files to create their own statistical tabulations and data summaries.

Specific microdata samples available on CD-ROM/DVD can be obtained through the census catalog available on the U.S. Census Bureau's home page (<http://www.census.gov/>).

#### SUMMARY TAPE FILES.

Summary File 1 (SF 1) contains 286 detailed tables focusing on age, sex, households, families, and housing units. These tables provide in-depth figures by race and Hispanic origin; some tables are repeated for each of nine race/Latino groups. Counts also are provided for over 40 American Indian and Alaska Native tribes and for groups within race categories. The race categories include 18 Asian groups and 12 Native Hawaiian and Other Pacific Islander groups. Counts of persons of Hispanic origin by country of origin (28 groups) are also shown.

Summary File 1 presents data for the United States, the 50 States, and the District of Columbia in a hierarchical sequence down to the block level for many tabulations, but only to the census tract level for others. Summaries are included for other geographic areas such as ZIP Code Tabulation Areas (ZCTAs) and Congressional districts.

Geographic coverage for Puerto Rico is comparable to the 50 States. Data are presented in a hierarchical sequence down to the block level for many tabulations, but only to the census tract level for others. Geographic areas include barrios, barrios-pueblo, subbarrios, places, census tracts, block groups, and blocks. Summaries also are included for other geographic areas such as ZCTAs.

Summary File 2 (SF 2) contains 47 detailed tables focusing on age, sex, households, families, and occupied housing units for the total population. These tables are repeated for 249 detailed population groups based on the following criteria:

- No tables are available for geographic areas having a population of less than 100.
- Tables are repeated only for the race groups, American Indian and Alaska Native tribes, and Hispanic or Latino groups having a population of 100 or more within the geographic area.

For a complete list of the 249 population groups, see Appendix H of the SF 2 Technical Documentation (PDF).

Summary File 3 consists of 813 detailed tables of Census 2000 social, economic, and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire. Fifty-one tables are repeated for nine major race and Hispanic or Latino groups: White alone; Black or African American alone; American Indian and Alaska Native alone; Asian alone; Native Hawaiian and Other Pacific Islander alone; Some other race alone; Two or more races; Hispanic or Latino; and White alone, not Hispanic or Latino.

Summary File 3 presents data for the United States, the 50 States, the District of Columbia, and Puerto Rico in a hierarchical sequence down to the block group for many tabulations, but only to the census tract levels for others. Summaries are included for other geographic areas such as Zip Code Tabulation Areas (ZCTAsTM) and Congressional districts (106th Congress).

Summary File 4 (SF4) contains the sample data, which is the information compiled from the questions asked of a sample of all people and housing units.

The sample data are presented in 213 population tables (matrices) and 110 housing tables, identified with "PCT" and "HCT," respectively. Each table is iterated for 336 population groups: the total population, 132 race groups, 78 American Indian and Alaska Native tribe categories (reflecting 39 individual tribes), 39 Hispanic or Latino groups, and 86 ancestry groups.

SF 4 is released as individual files for each of the 50 States, the District of Columbia, and Puerto Rico; and for the United States. The tables (matrices) are identical for all files, but the geographic coverage differs. Data are provided down to the census tract level.

Reports: See <http://www.census.gov/main/www/cen2000.html>.

Future Plans: The next decennial census will be conducted in 2010. Reengineering of the 2010 census includes replacing the long form with the ACS. The ACS is a nationwide survey designed to provide communities a fresh look at how they are changing. It is intended to eliminate the need for the long form in the 2010 Census. The ACS collects information from U.S. households and group quarters similar to what was collected on the Census 2000 long form, such as income, commute time to work, home value, veteran status, and other important data. As with the official U.S. census, information about individuals will remain confidential.

For more information:

Email: [pio@census.gov](mailto:pio@census.gov)

Phone: 301-763-3691

Website: <http://www.census.gov/main/www/cen2000.html>

## **International Data Base (IDB)**

Sponsor: U.S. Census Bureau

Data Collection Agency/Organization: U.S. Census Bureau, International Programs Center

Description: The International Data Base (IDB) is a computerized source of demographic and socioeconomic statistics for 227 countries and areas of the world. The IDB was created in the U.S. Census Bureau's International Programs Center (IPC) in response to the information requirements of IPC staff to meet the needs of organizations that sponsor IPC's research efforts. The IDB provides quick access to specialized information, with emphasis on demographic measures, for individual countries or selected groups of countries of the world.

The IDB combines data from country sources (especially censuses and surveys) with IPC's estimates and projections to provide information dating back as far as 1950 and as far ahead as 2050. Because the IDB is maintained at IPC as a research tool in response to sponsor requirements, the amount of information available for each country may vary. As funding and research activity permit, IPC will update and expand the coverage. The various government agencies, international organizations, and others who sponsor IPC's research program support and benefit from the IDB.

The major types of data available in the IDB include:

- Population by age and sex
- Vital rates, infant mortality, and life tables
- Fertility and child survivorship
- Migration
- Marital status
- Family planning
- Ethnicity, religion, and language
- Literacy
- Labor force, employment, and income
- Households

Data characteristics:

- Temporal: Selected years, 1950–present, projected demographic data to 2050.
- Spatial: 227 countries and areas.
- Resolution: National population, selected data by urban/rural residence, selected data by age and sex.

Sources of the data include:

- U.S. Census Bureau, Estimates and Projections
- National Statistics Offices
- United Nations and Specialized Agencies (ILO, UNESCO, WHO)

For more information:

Contact: Pat Dickerson or Peter Johnson

Email: [pop@census.gov](mailto:pop@census.gov)

Phone: 301-763-2422

Website: <http://www.census.gov/ipc/www/idb/>

## **Longitudinal Employer-Household Dynamics (LEHD)**

Sponsor: National Institute on Aging, U.S. Census Bureau, National Science Foundation

Data Collection Agency/Organization: U.S. Census Bureau

**Purpose:** The mission of LEHD is to combine Federal and State administrative data on employers and employees with core Census Bureau censuses and surveys to improve the quality and understanding of survey products, and to conduct or facilitate research on emerging social and economic policy issues. Of particular importance is the development of a data infrastructure of integrated household and firm datasets that relate employers to their employees and vice-versa. This data infrastructure facilitates longitudinal research applications in both the household/individual and firm/establishment dimensions. The specific research is targeted at filling an important gap in the available data on older workers by providing information on the demand side of the labor market.

**Description:** These datasets comprise Title 13 protected data from the Current Population Surveys, Surveys of Income and Program Participation, American Community Surveys, the Business Register, and Economic Censuses and Surveys. We have built employer-employee data relations based on universe State Unemployment Insurance wage records and Quarterly Census of Employment and Wages from all but a few states. LEHD's second method of developing employer-employee data relations through the use of Federal tax data has been completed.

LEHD has achieved its objective by:

- Producing summary tables on accessions, separation, job creation, destruction and earnings by age and sex of worker — by industry and geographic area. These tables are available for 32 States and on the website at <http://lehd.dsd.census.gov>. Additional web-based analytical tools are also available on the website.
- Providing State-level micro data to approved researchers at Census Bureau facilities. The data files consist of longitudinal datasets on all firms in each participating State (quarterly data, 1991–2005), with information on age, sex, turnover, and skill level of the workforce as well as standard information on employment, payroll, sales, and location.
- Creating new public use files that are derived from the integration of the Survey of Income and Program Participation with detailed earnings histories. A first version of the inference valid synthetic data files has been made available to SSA, CBO and Census, and is being validated.
- Creating new data products (summary tables and enhanced public use data files) by linking worker information to the Business Register that is enhanced by detailed and universal firm pension and benefit information from the 5500 file. This file provides information on whether firms cover employees on a plethora of pension plans, including profit-sharing, 401(k)s, 403(b)s, IRA's, and defined contribution plans. This work is reported as LEHD Technical Paper No. TP-2003-05, "Describing the Form 4400-Business Register Match."

New data products include: Current Population Survey (1973–2003) enhanced by detailed information on pension plans offered by previous and current employers, the firm take-up rate, and financial characteristics of the firm; Survey of Income and Program Participation (1984, 1990–93, 1996; 2001; 2004) enhanced by detailed information on pension plans offered by previous and current employers, the firm take-up rate, and financial characteristics of the firm; and summary information on pension plan coverage by firm type and workforce characteristics from the Business Register.

Data Availability: Research conducted on the LEHD data and other products developed under this proposal at the Census Bureau takes place under a set of rules and limitations that are considerably more constraining than those prevailing in typical research environments. Successful peer-reviewed proposals to carry out research using the LEHD data must be approved by the Census Bureau, which will confirm that:

- the Bureau would benefit from the proposed research (a requirement for granting access to Title 13 protected data);
- the proposed research is feasible using the LEHD data;
- the variables and cases requested can be supplied; and
- appropriate resources exist to undertake the analysis.

If State data are requested, the successful peer-reviewed proposals must also be approved by the participating State. If Federal tax data are requested, the successful peer-reviewed proposals must also be approved by the Internal Revenue Service. Researchers using the LEHD data will be required to obtain Special Sworn Status from the Census Bureau and be subject to the same legal penalties as regular Census Bureau employees for disclosure of confidential information.

For more information:

Contact: Jeremy Wu

Email: [Jeremy.s.wu@census.gov](mailto:Jeremy.s.wu@census.gov)

Phone: 301-763-5290

Website: <http://lehd.did.census.gov>

## **National Longitudinal Mortality Study (NLMS)**

Sponsors: National Heart, Lung, and Blood Institute; National Cancer Institute; National Institute on Aging; National Center for Health Statistics; U.S. Census Bureau

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To study the effects of demographic and socioeconomic characteristics on differentials in U.S. mortality rates.

Study Design: The NLMS currently consists of Annual Social and Economic Supplements, which cover the period from March 1973 to March 2002, Current Population Surveys for February 1978, April 1980, August 1980, December 1980, and September 1985, and one 1980 Census cohort, 26 cohorts in all. Mortality information is obtained from death certificates available for deceased persons through the National Center for Health Statistics. Important variables available for analyses are standard demographic and socio-economic variables such as, education, income, and employment as well as information collected from death certificates, including cause of death.

Data Limitations: The NLMS is based on specific survey months of the Current Population Survey, the Annual Social and Economic Supplement, and a subset of the 1980 Census. These are one-time data collection processes with no subsequent data collection. Therefore, one limitation of NLMS data is that they provide a one-time only baseline measurement of subjects in a long-term follow-up situation. Another limitation of these data is that, although the CPS and census instruments do provide extensive data collection capabilities in specific subject matter areas, desirable general or specific health information is not collected, and smoking status is available on only a limited number of records.

Data Availability: An NLMS public-use file is available to interested researchers upon request through NLMS-Census Bureau principal investigator, Norm Johnson. Due to the confidential nature of the NLMS Title 13 data, the NLMS Public-use file consists of a restricted set of NLMS variables.

Research access to the entire NLMS database may be arranged through the principal investigators of the NLMS sponsoring agencies. Research topics of interest should fall within the general health interests of the agency. Principal investigators sponsor research through a quick-turnaround approval process established by the NLMS Steering Committee. The committee determines the priority of all NLMS research. Committee approved projects are assigned to an NLMS statistician who works directly with the researcher as a statistical consultant and as the interface to the NLMS database. Results are delivered to researchers on a, short-turnaround, flow basis through the most convenient means available in either electronic or hard copy format.

Researchers may also be interested in gaining access to a file containing the same information as the NLMS public-use file but which also includes geographical information. This file is available through the Census Bureau's Research Data Centers. For complete information on the



use of the Census Bureau's Research Data Centers go to the Center for Economic Studies Census Bureau's website: <http://www.ces.census.gov/>.

A researcher may work directly with an NLMS statistician on-site at the Census Bureau by becoming a Special Sworn Census Bureau employee. A person interested in this option must first gain research sponsorship of a project through an NLMS sponsoring agency according to the NLMS Steering Committee approval process and then apply for Special Sworn status through Norm Johnson, Census Bureau NLMS Principal Investigator.

For more information:

Contact: Norman J. Johnson

Email: [norman.j.johnson@census.gov](mailto:norman.j.johnson@census.gov)

Phone: 301-763-4270

Website: [www.census.gov/nlms/](http://www.census.gov/nlms/)

## **Population Projections**

Sponsor: U.S. Census Bureau

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: Information about the possible future race/origin/age/sex composition of the United States.

Research Design: The population projections for the United States are interim projections that take into account the results of Census 2000. These interim projections were created using the cohort-component method, which uses assumptions about the components of population change. They are based on Census 2000 results, official post-census estimates, as well as vital registration data from the National Center for Health Statistics. The assumptions are based on those used in the projections released in 2000 that used a 1998 population estimate base. Some modifications were made to the assumptions so that projected values were consistent with estimates from 2001 as well as Census 2000.

Fertility is assumed to increase slightly from current estimates. The projected total fertility rate in 2025 is 2.180, and it is projected to increase to 2.186 by 2050. Mortality is assumed to continue to improve over time. By 2050, life expectancy at birth is assumed to increase to 81.2 for men and 86.7 for women. Net immigration is assumed to be 996,000 in 2025 and 1,097,000 in 2050.

Race and Hispanic origin: Interim projections based on Census 2000 were also done by race and Hispanic origin. The basic assumptions by race used in the previous projections were adapted to reflect the Census 2000 race definitions and results. Projections were developed for the following groups: (1) non-Hispanic white alone, (2) Hispanic white alone, (3) black alone, (4) Asian alone, and (5) all other groups. The fifth category includes the categories of American Indian and Alaska Native, Native Hawaiian and Other Pacific Islanders, and all people reporting more than one of the major race categories defined by the Office of Management and Budget (OMB).

For a more detailed discussion of the cohort-component method and the assumptions about the components of population change, see U.S. Census Bureau, Population Division Working Paper No. 38, "Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100," by Hollmann, Mulder, and Kallan. While this paper does not incorporate the updated assumptions made for the interim projections, it provides a more extensive treatment of the earlier projections, released in 2000, on which the interim series is based.

For more information:

Contact: Population Projections Branch

Phone: 301-763-2428

Website: <http://www.census.gov/population/www/projections/index.html>

## **Survey of Income and Program Participation (SIPP)**

Sponsor: U.S. Census Bureau, Social Security Administration

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To collect source and amount of income, labor force information, program participation and eligibility data, and general demographic characteristics to measure the effectiveness of existing Federal, State, and local programs; to estimate future costs and coverage for government programs, such as food stamps; and to provide improved statistics on the distribution of income in the country. SIPP also offers detailed information on cash and noncash income on a sub-annual basis in addition to collecting data on taxes, assets, liabilities, and participation in government transfer programs.

Survey Universe: U.S. civilian noninstitutionalized population

Research Design: This is a longitudinal survey—a continuous series of national panels.

Survey Mode: Most interviews conducted through 1991 were in the form of personal visits. In 1992, SIPP switched to maximum telephone interviewing to reduce costs. Wave 1, 2, and 6 interviews were still conducted in person, but other interviews were conducted by telephone to the extent possible. SIPP telephone interviews and personal visits are carried out by the same interviewer interacting with the same respondents. Interviewers typically make phone calls from their homes. For security and confidentiality reasons, they are not allowed to use cellular or cordless telephones in the interviews. If a standard telephone is not available, the interviews must be conducted face-to-face. Repeated failure to reach a respondent by telephone may also require an in-person visit to the listed address.

Unit of Analysis: All household members 15 years old and over are interviewed by self-response, if possible; proxy response is permitted when household members are not available for interviewing.

Sample: The SIPP sample is a multistage-stratified sample of the U.S. civilian noninstitutionalized population with sample sizes ranging from approximately 14,000 to 36,700 interviewed households per panel. The duration of each panel ranges from 2½ years to 4 years.

The Census Bureau also over sampled the low-income population for the 1996, 2001, and 2004 Panels using decennial census information. Housing units within each PSU were split into high- and low-poverty strata. If the housing unit received the Census long form that included income questions, the unit's poverty status was determined directly; for other housing units, poverty status was assumed on the basis of responses to Census short-form items predictive of poverty rates.

Topics: Income, labor force participation, program participation, and eligibility.

Data Availability: For the 1984–93 panels, a panel of households was introduced each year in February. A 4-year panel was introduced in April 1996. A 2000 panel was introduced in February 2000 for 2 waves. A 3-year 2001 panel was introduced in February 2001, and a 2½ year 2004 panel was introduced in February 2004.

Linked Data: The Census SIPPs are linked to the IRS wage records and 1040SEs and 1040 Tips. These exist for each year from 1982 to most recent year lagged one year. The Social Security earnings are from IRS forms and owned by IRS. They are the employer reported wages and salaries and self-employed reported income subject to taxation by Social Security up to the maximum subject to tax. They exist for each year from 1951 to most recent year lagged one year. The most recent year lagged one year is currently 2004.

SSA benefit records contain information on Medicare Part A, Part B, and Part D low income subsidy and Medicaid subsidies of Medicare Part B (QMB, SLMB, QI).

Data Dissemination: Data collected in SIPP and supporting documentation are available in various forms. They include published estimates based on those data, micro data in several formats, documentation for each of the micro data files, and more general documentation about methodological issues in SIPP. The latter includes the SIPP Quality Profile, a series of working papers distributed by the Census Bureau, articles published in academic journals, and conference proceedings.

SIPP micro data files can be obtained from several sources. All public use micro data files can be obtained on CD-ROM directly from the Census Bureau. SIPP micro data are available online from the SIPP website at <http://www.census.gov/sipp/>. The Internet site offers two data access tools DataFerrett and the SIPP FTP site, DataFerrett is a system that enables users to access and manipulate large demographic and economic data sets on-line. The SIPP FTP site has data files and documentation for downloading.

Cross-sectional data are presented for various socioeconomic characteristics for a 4-month period. Longitudinal data are presented for a 2½-year or 3-year period. Variables for both data sets include age, race, sex, Hispanic origin, marital status, household/family relationship, educational attainment, work experience, and income. Basic cross-sectional questions are supplemented with topically relevant questions such as employment history, work disability, education, health care, financial assets, retirement accounts, etc.

Reports: SIPP publications can be found at [www.sipp.census.gov/sipp/pubs.html](http://www.sipp.census.gov/sipp/pubs.html).

Future Plans: SIPP has been an important source of data reflecting the economic well-being of U.S. households over time. However, the length and breadth of the interviews and the longitudinal household design have resulted in ever-increasing number of refusals to participate. Many steps have been taken to curtail attrition, including monetary incentives, but have not been successful in reducing it to an acceptable level. In addition, the complexity of the instrument, has led to long delays before the data can be understood, documented, and finally, disseminated.

In FY2007, the Census Bureau will begin planning and development for a new approach to providing wealth, income, health insurance, and program participation data for the United States. These plans will take advantage of the advances that the Census Bureau has made in acquiring and integrating administrative records with survey data in recent years and in modeling for local area estimates.

This improved measurement program will be a collaborative effort among the traditional SIPP stakeholders, such as the Office of Management and Budget, the Social Security Administration, the Department of Health and Human Services, the Department of Agriculture, the Department of Labor, the Small Business Administration, and the Congressional Budget Office. The focus will be on meeting the information needs of policymakers, these agencies, and the research community.

The Census Bureau believes that its existing surveys will provide the baseline data from which it can obtain current and retrospective income and program data from administrative records and then conduct follow-on interviews to obtain more detailed socioeconomic data unavailable elsewhere. Most likely, this new program will include an annual survey that will probe issues such as health, disability, retirement expectations, and job search that will lead to an understanding of the life events that lead people to enter or leave government assistance programs.

The Census Bureau envisions a less burdensome, less expensive world-class data system that measures income, wealth, and program dynamics.

For more information:

Email: [hhes.sipp.survey@census.gov](mailto:hhes.sipp.survey@census.gov)

Website: [www.sipp.census.gov/sipp/](http://www.sipp.census.gov/sipp/)

**Department of Health and Human Services  
Administration on Aging**

**Census 2000 Special Tabulation on Aging**

Sponsor: Administration on Aging

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To provide State and area agencies on aging with detailed small-area data from Census 2000 on the number and characteristics of the over (60+) population.

The Special Tabulation on Aging (STA) is a set of 116 population tables and 69 housing and household tables that were tabulated by the Census Bureau for AoA from the Census 2000 sample (“long form”) data. The topics covered in these tabulations include age, sex, race, Hispanic or Latino origin, marital status, income, poverty, disability, and much more. This tabulation provides much more detail on the older population (60+) than is available from other Census sources.

The data are organized by individual tables. Each table includes data for a State (including the District of Columbia and Puerto Rico) and data are shown for a number of geographical levels: United States (50 States +DC), State, Planning and Service Area (PSA—the geographic area served by a single area agency on aging), county, county subdivision in 12 States with a population of 2,500 or more, place with a population of 2,500 or more, census tract, as well as American Indian and Alaska Native areas. Also, the urban and rural components of States and PSAs are shown.

The data of the Special Tabulation on Aging along with documentation are available on the AoA website at <http://www.aoa.gov/prof/Statistics/Tab/specialtab.aspx>.

For more information:

Contact: Saadia Greenberg

Email: [saadia.greenberg@aoa.gov](mailto:saadia.greenberg@aoa.gov)

Phone: 202-357-3554

Website: <http://www.aoa.gov/prof/Statistics/Tab/specialtab.aspx>

## Agency for Healthcare Research and Quality

### Healthcare Cost and Utilization Project (HCUP)

Sponsor: Agency for Healthcare Research and Quality, Center for Delivery, Organization and Markets

Data Collection Agency/Organization: A Federal-State-Industry partnership sponsored by the Agency for Healthcare Research and Quality

Purpose: HCUP is a federally-funded project that builds on the efforts of State data organizations, hospital associations, and private data organizations to create a national resource of health care data. HCUP maintains the largest collection of longitudinal, all payer hospital care databases in the United States, enabling research on a broad range of health policy issues.

Description: The Healthcare Cost and Utilization Project (HCUP, pronounced “H-Cup”) ([www.hcup-us.ahrq.gov/home.jsp](http://www.hcup-us.ahrq.gov/home.jsp)) is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by AHRQ. HCUP databases are made possible through voluntary efforts of State data organizations, hospital associations, and private data organizations to create a national information resource of encounter-level health care data. HCUP administrative data includes all-payer categories (including the uninsured), and goes back to 1988 for some datasets. The databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, patient safety, access to health care programs, and outcomes of treatments at the national, State, and local market levels.

Because of their large size and scope, HCUP databases enable unique analyses: they include information on specific medical conditions and procedures (including rare events); they are useful in tracking utilization for population subgroups, such as minorities, children, women, and the uninsured; and they can be employed to analyze different geographic levels (national, regional, State, and community) within the United States. To augment the HCUP databases, AHRQ has developed publicly available tools and software and Web-based products that can be used by audiences with varying levels of research experience.

HCUP databases are available for purchase through the HCUP Central Distributor ([www.hcup-us.ahrq.gov/tech\\_assist/centdist.jsp](http://www.hcup-us.ahrq.gov/tech_assist/centdist.jsp)). Prospective purchasers and all persons with access to the databases are required to read and sign a Data Use Agreement.

The following files are distributed through the HCUP Central Distributor:

The **Nationwide Inpatient Sample (NIS)** ([www.hcup-us.ahrq.gov/nisoverview.jsp](http://www.hcup-us.ahrq.gov/nisoverview.jsp)) is a database of hospital inpatient stays. Researchers and policymakers use the NIS to identify, track, and analyze national trends in health care utilization, access, charges, quality, and outcomes. It is the largest all-payer inpatient care database in the United States containing data from approximately 8 million hospital stays per year. The sampling frame for the NIS is a sample of hospitals that

comprises approximately 90 percent of all hospital discharges in the United States. NIS data are produced annually, and are currently available from 1988 to 2006, allowing analysis of trends over time. The number of States in the NIS has grown from 8 in the first year to 38 in the current year. The NIS is the only national hospital database with charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured. Its large sample size enables analyses of rare conditions, such as congenital anomalies; uncommon treatments, such as organ transplantation; and special patient populations, such as children. NIS includes hospital identifiers that permit linkages to the American Hospital Association's Annual Survey and county identifiers that permit linkages to the Area Resource File.

The **Kids' Inpatient Database (KID)** ([www.hcup-us.ahrq.gov/kidoverview.jsp](http://www.hcup-us.ahrq.gov/kidoverview.jsp)) is a nationwide sample of pediatric inpatient discharges and was specifically designed to permit researchers to study a broad range of conditions and procedures related to child health issues. Produced every 3 years since the 1997 data year, the most recent KID is for 2006 and contains data drawn from 38 State Inpatient Databases and includes 2–3 million hospital discharges for children 0-20 years of age. The sample of pediatric discharges comes from between 2,500 to 4,000 U.S. community hospitals (defined as short-term, non-Federal, general and specialty hospitals, excluding hospital units of other institutions). The KID's large sample size enables analyses of both common and rare conditions such as congenital anomalies, uncommon treatments, and organ transplantation. The data includes charge information on all patients, including children covered by Medicaid, private insurance, and the uninsured.

The **State Inpatient Databases (SID)** ([www.hcup-us.ahrq.gov/sidoverview.jsp](http://www.hcup-us.ahrq.gov/sidoverview.jsp)) are produced annually and contain the universe of inpatient discharge abstracts from the currently 39 participating States. The data are transformed into a uniform format to facilitate multi-State comparisons and analyses. Together, the SID encompasses about 90 percent of all U.S. community hospital discharges. Some States include discharges from specialty facilities, such as acute psychiatric hospitals. The SID contain a core set of clinical and non-clinical information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured. In addition to the core set of uniform data elements common to all SID, some include other elements, such as the patient's race.

The **State Ambulatory Surgery Databases (SASD)** ([www.hcup-us.ahrq.gov/sasdooverview.jsp](http://www.hcup-us.ahrq.gov/sasdooverview.jsp)) are produced annually and contain data from ambulatory care encounters from hospital-affiliated and sometimes freestanding ambulatory surgery sites. The SASD databases capture surgeries performed on the same day in which patients are admitted and discharged. The data represent ambulatory surgery encounters in non-inpatient health care settings such as ambulatory surgery centers, freestanding clinics, and hospital outpatient departments from States selected by AHRQ. Currently 27 States participate in the SASD.

The **State Emergency Department Databases (SEDD)** ([www.hcup-us.ahrq.gov/seddoverview.jsp](http://www.hcup-us.ahrq.gov/seddoverview.jsp)) are produced annually and contain data from hospital-affiliated emergency departments for visits that do not result in hospitalizations. The SEDD are designed to allow analyses of such emergency department-related concerns as patient flow, access to care, and practice variations in emergency department care settings. Currently 25 States participate in the SEDD.



## HCUP Tools and Software Products:

**HCUP Tools and Software** can be applied to HCUP data or other administrative data. Popular tools include methods for classifying ICD-9-CM diagnoses or procedures into clinically meaningful categories, indicators for chronic conditions and cost-to charge conversion ratios. HCUP tools and software are described on the HCUP User Support website ([www.hcup-us.ahrq.gov/tools\\_software.jsp](http://www.hcup-us.ahrq.gov/tools_software.jsp)).

Two widely used tools are HCUPnet and the AHRQ Quality Indicators (QIs).

**HCUPnet** ([www.hcupnet.ahrq.gov](http://www.hcupnet.ahrq.gov)) is a free, online interactive tool for identifying, tracking, analyzing, and comparing statistics on hospital care. HCUPnet queries generate statistics in a table format using data from the NIS, SID and SEDD databases for those States that have agreed to participate. Using HCUPnet's easy step-by-step query system, you can generate tables and graphs on national and regional statistics and trends for community hospitals in the United States. In addition, community hospital data are available for those States that have agreed to participate in HCUPnet.

HCUPnet also provides you with statistics based on the AHRQ Quality Indicators (QIs) that have been applied to the HCUP Nationwide Inpatient Sample. These statistics provide insight into potential quality of care problems. You can download software for the QIs from the Quality Indicators Website ([www.qualityindicators.ahrq.gov](http://www.qualityindicators.ahrq.gov)) and apply them to your own data. The AHRQ Quality Indicators (QIs) are measures of health care quality that make use of readily available hospital inpatient administrative data. Software and user guides for all modules are available to assist users in applying the Quality Indicators to their own data.

The **AHRQ QIs** consist of five modules measuring various aspects of quality:

- Prevention Quality Indicators—Identify hospital admissions that evidence suggests could have been avoided, at least in part, through high-quality outpatient care.
- Inpatient Quality Indicators—Reflect quality of care inside hospitals including inpatient mortality for medical conditions and surgical procedures.
- Patient Safety Indicators—Also reflect quality of care inside hospitals, but focus on potentially avoidable complications and iatrogenic events.
- Pediatric QIs—Reflect both quality of care inside hospitals and identify potentially avoidable hospitalizations among children.
- Neonatal QIs--Reflect the quality of care inside hospitals for neonates 0-28 days old with a birth weight between 500 and 1499 grams or a gestational age between 24-30 weeks.

## Print Products:

### HCUP Fact Books:

- Care of Adults with Mental Health and Substance Abuse Disorders in U.S. Community Hospitals, 2004 ([www.ahrq.gov/data/hcup/factbk10](http://www.ahrq.gov/data/hcup/factbk10))
- Ambulatory Surgery in U.S. Hospitals, 2003 ([www.ahrq.gov/data/hcup/factbk9](http://www.ahrq.gov/data/hcup/factbk9))

- Serving the Uninsured: Safety-Net Hospitals, 2003 ([www.ahrq.gov/data/hcup/factbk8](http://www.ahrq.gov/data/hcup/factbk8))
- Procedures in U.S. Hospitals, 2003 ([www.ahrq.gov/data/hcup/factbk7](http://www.ahrq.gov/data/hcup/factbk7))
- Hospitalization in the United States, 2002 ([www.ahrq.gov/data/hcup/factbk6](http://www.ahrq.gov/data/hcup/factbk6))
- Preventable Hospitalizations, 2000 ([www.ahrq.gov/data/hcup/factbk5](http://www.ahrq.gov/data/hcup/factbk5))
- Care of Children and Adolescents in U.S. Hospitals ([www.ahrq.gov/data/hcup/factbk4](http://www.ahrq.gov/data/hcup/factbk4))
- Care of Women in U.S. Hospitals 2000 ([www.ahrq.gov/data/hcup/factbk3](http://www.ahrq.gov/data/hcup/factbk3))
- Procedures in U.S. Hospitals, 1997 ([www.ahrq.gov/data/hcup/factbk2](http://www.ahrq.gov/data/hcup/factbk2))
- Hospitalization in the United States, 1997 ([www.ahrq.gov/data/hcup/factbk1](http://www.ahrq.gov/data/hcup/factbk1))

### **Methodology Reports:**

HCUP Methods Series features a broad array of methodological information on the HCUP databases and software tools. Reports in the HCUP Methods Series ([www.hcup-us.ahrq.gov/reports/methods.jsp](http://www.hcup-us.ahrq.gov/reports/methods.jsp)) are listed in chronological order.

### **HCUP Database Reports:**

HCUP Database Reports ([www.hcup-us.ahrq.gov/reports.jsp](http://www.hcup-us.ahrq.gov/reports.jsp)) are specific to the design and use of the HCUP databases.

- Nationwide Inpatient Sample (NIS) reports
- Kids' Inpatient Database (KID) reports
- State Inpatient Databases (SID) reports
- State Ambulatory Surgery Databases (SASD) reports
- State Emergency Department Databases (SEDD) reports

### **HCUP Statistical Briefs:**

HCUP Statistical Briefs present simple, descriptive statistics on a variety of specific, focused topics. New topics are released approximately every two weeks. The Statistical Brief reports are listed in reverse chronological order on the HCUP user support website ([www.hcup-us.ahrq.gov/reports/statbriefs.jsp](http://www.hcup-us.ahrq.gov/reports/statbriefs.jsp)).

### **HCUP Highlights and Other Publications:**

HCUP Highlights illustrate key findings from the HCUP databases.

- Economic and Health Costs of Diabetes ([www.ahrq.gov/data/hcup/highlight1/high1.htm](http://www.ahrq.gov/data/hcup/highlight1/high1.htm))
- Hospital and Ambulatory Surgery Care for Women's Cancers ([www.ahrq.gov/data/hcup/highlight2/high2.htm](http://www.ahrq.gov/data/hcup/highlight2/high2.htm))
- The Value of Hospital Discharge Data ([www.hcup-us.ahrq.gov/reports/final\\_report.pdf](http://www.hcup-us.ahrq.gov/reports/final_report.pdf))

### **HCUP Facts and Figures Reports:**

HCUP Facts and Figures provides an overview of national statistics on hospital stays and trends.

- Statistics on hospital-based care in the United States, 2006 ([www.hcup-us.ahrq.gov/reports/factsandfigures/HAR\\_2006.pdf](http://www.hcup-us.ahrq.gov/reports/factsandfigures/HAR_2006.pdf))

- Statistics on hospital-based care in the United States, 2006  
([www.hcup-us.ahrq.gov/reports/factsandfigures/HAR\\_2005.pdf](http://www.hcup-us.ahrq.gov/reports/factsandfigures/HAR_2005.pdf))

**Publications based on HCUP data** are available in the HCUP section of the AHRQ website ([www.ahrq.gov/data/hcup](http://www.ahrq.gov/data/hcup)) (go to “Publications from the HCUP Databases” under “New Findings and Publications”). Publications are listed by author. Information includes title, publication, data, and sometimes access to an abstract.

### **Coming Soon:**

HCUP is continually evolving and expanding its offerings. Check the HCUP User Support Website to find What’s New ([www.hcup-us.ahrq.gov/](http://www.hcup-us.ahrq.gov/)). Two important features planned for the near future are:

The **Nationwide Emergency Department Sample (NEDS)**, the first and only all-payer emergency department (ED) data source in the U.S. providing national estimates of ED encounters. The NEDS will be the newest addition to the family of HCUP databases. It can be used to create national estimates of ED visits – both treat-and-release and those that resulted in a hospital admission. It contains clinical and non-clinical information on all patients, regardless of payer – including Medicare, Medicaid, private insurance and the uninsured (expected January 2009).

The **Efficiency and Quality Improvement Portal (EQUIP)**, a tool which will allow local users (e.g., statewide data organizations) to apply AHRQ tools to their own data and then generate a website which they will host on their own server.

The initial version of the tool will generate utilization, cost, and quality statistics using hospital administrative data and is based on HCUPnet and the AHRQ Quality Indicators. Local host-users will be able to input their own data, step through the process of building the files needed for the system, customize a prototype website, generate a website they will host, and make the information available at whatever level they choose, such as at the hospital and area level.

### **For more information about HCUP:**

Contact: HCUP User Support

Email: [hcup@ahrq.gov](mailto:hcup@ahrq.gov)

Phone: 1-866-290-HCUP

Website: [www.ahrq.gov/data/hcup/](http://www.ahrq.gov/data/hcup/); HCUP User Support Website: [www.hcup-us.ahrq.gov/](http://www.hcup-us.ahrq.gov/)

## **Medical Expenditure Panel Survey (MEPS)**

Sponsors: Agency for Healthcare Research and Quality, National Center for Health Statistics

Data Collection Agency/Organization: Westat, U.S. Census Bureau

Purpose: The MEPS is conducted to provide nationally representative estimates of health care use, health care expenditures, sources of payment, health insurance coverage and health status for the U.S. civilian noninstitutionalized population. The MEPS is comprised of three component surveys: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC). The Household Component is the core survey, and it forms the basis for the MPC sample. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications. The MEPS was initiated in 1996 and is a continuous ongoing survey. MEPS predecessor surveys were conducted in 1987 and 1977.

The 1996 MEPS also included a nationally representative sample of nursing homes and persons who were nursing home residents at any time during 1996. Details on the Nursing Home Component (NHC) are provided on the MEPS website.

Survey Universe: U.S. civilian noninstitutionalized population.

Description: The MEPS HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income and assets, retirement, and current employment. The HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a 2½-year period. Using computer-assisted personal interviewing technology, data on medical expenditures and use for two calendar years are collected on all household members. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures. Data can be analyzed at the person level, the event level, the family level, or the health insurance unit level. Each annual sample size is about 15,000 households; the sample size of persons age 65 and over is approximately 4,000 (for the 2001 MEPS and in subsequent years). The full-year household core response rate has generally ranged from about 65 to 71 percent.

A number of quality-related enhancements were made to the MEPS beginning in 2000, including the fielding of an annual adult self-administered questionnaire (SAQ). This questionnaire contains items on patient satisfaction and accountability measures from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®; previously known as the Consumer Assessment of Health Plans), the SF-12 physical and mental health assessment tool, EQ-5D EuroQol 5 dimensions with visual scale (2000–03), and several attitude items. Starting in 2004

the K-6 Kessler mental health distress scale and the PH2 two-item depression scale were added to the SAQ.

**Description:** The MEPS Insurance Component (IC) is an annual panel survey that collects data on health insurance plans obtained through employers, unions, and other sources of private health insurance. Information on premiums, deductible and co-payment provisions, distribution of premium and costs across payers are obtained. Data are collected by the Census Bureau from the sampled organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone follow-up for nonrespondents. The MEPS IC supports estimation at the national and State level.

**Description:** The primary purpose of the Medical Provider Component (MPC) is to collect detailed charge and payment data from hospitals, physicians, home health care providers, and pharmacies to supplement/replace information received from MEPS HC respondents to aid in the estimation of health care expenses. The MPC is conducted through telephone interviews and mailed survey materials.

**Data Availability:** MEPS HC data releases, including documentation and codebooks, are available free to the public on the Internet (via the MEPS website). A series of calendar-specific MEPS public use data files are produced annually. Each of these files includes full-year information from several rounds of data collection, which together comprise a complete calendar year's worth of information. Full Year data files vary in structure depending on the nature of file content. Files are produced at the person, event, condition, job, and person-round level. MEPS also releases Point-In-Time Files (a snap shot of what is going on at a fixed point in time) during the first part of the calendar year. These files contain minimum data elements and are intended to give data users/analysts an early glimpse of how the full year estimates will look. Special topic files are based on data collected in supplements to the MEPS on intermittent bases. For a chart listing the MEPS supplements and the rounds in which they were collected, see the "Data Collection Schedule for MEPS HC Supplements" on the MEPS website.

MEPS IC data are published in tabular format on a yearly schedule. Access to the MEPS Insurance Component data is only available in a Census Bureau Research Data Center.

A series of customizable pre-defined tables are also available for each data year in the MEPS Tables Compendia section of the MEPS website.

MEPS data (HC and IC data) are also available via MEPSnet, an on-line, interactive, statistical tool developed to give users the ability to analyze MEPS data in real-time.

Many of the MEPS databases include considerably more data than can be made available to the general public because of the constraints of our confidentiality guidelines. In order to facilitate the use of such data, while maintaining the confidentiality, AHRQ has developed a Data Center (a physical space at AHRQ in Rockville, Maryland) where researchers with approved projects can access data files not available for public dissemination. Data available for use in the AHRQ Data Center include the 1996 MEPS Nursing Home Component, selected data from the MEPS

Medical Provider Component, and analyses of MEPS HC data linked to secondary data sources. See the MEPS website for details.

**Linked Data:** For approved Data Center projects MEPS data can be linked to secondary data files at the State, county, and Zip code level, including linkage to the Area Resources File. For 1987 and 1996 analytic files that include marginal tax rates are available. For 1996–99 linked HC-IC data are available. Researchers with approved Data Center projects can also supply their own data to be merged with AHRQ MEPS data.

A file containing a cross-walk to link each MEPS panel to the previous years National Health Interview Survey Public Use File is available upon request from AHRQ.

**Data Dissemination:** The primary method of MEPS data dissemination is via the MEPS website. A variety of print and electronic products are generated from the data collected through MEPS: public use data files and statistical tables; periodic data reports (Research Findings and Chartbooks); short data summaries (Highlights, Facts and Figures, and web-based Statistical Briefs); and other research publications (Methodology Reports and Journal Articles). MEPS public use data files are produced annually and are released as data files in Portable Document Format (PDF), HyperText Markup Language (HTML), ASCII format or as data files in ASCII format containing SAS programming Statements. Files released for data years 2001 and beyond also contain SPSS format Statements. Some files may be compressed (“.exe” extension) or “zipped” (“.zip” extension) to reduce downloading time.

**Reports:** For copies of data products and reports, see the MEPS website. Selected MEPS data products are available from the AHRQ Publications Clearinghouse, phone: 1-800-358-9295; outside the United States: 703-437-2078; TDD for the hearing impaired, toll free: 888-586-6340. Address: P.O. Box 8547, Silver Spring, MD 20907. A sample of MEPS publications is provided below:

Banthin, J. and D. Bernard, PhD. Out-of-Pocket Expenditures on Health Care and Insurance Premiums among the Elderly Population, 2003. Rockville (MD): Agency for Healthcare, Research, and Quality, 2006. Statistical Brief #122. Web only publication.

Bernard, D., Banthin, J., and Encinosa, W. Health Care Expenditure Burdens Among Adults With Diabetes in 2001: *Medical Care* 2006; March; 44(3):210-215.

Cohen, J. Design and methods of the Medical Expenditure Panel Survey Household Component. Rockville (MD): Agency for Health Care Policy and Research, 1997. MEPS Methodology Report No. 1. AHCPR Pub. No. 97-0026.

Cohen, S., and Buchmueller, T. Trends in medical care costs, coverage, use, and access. *Medical Care* 2006; May; (5 Suppl):I1-I-3.

Fleishman, J.A. Demographic and Clinical Variations in Health Status. Rockville (MD): Agency for Healthcare, Research, and Quality, 2005. Methodology Report #15. Web only publication.

Miller, G. E. and Stagnitti, M. N. Trends in Statin Use in the Civilian Noninstitutionalized Medicare Population, 1997 and 2002. Statistical Brief #97. September 2005. Agency for Healthcare Research and Quality, Rockville, MD.

<http://www.meps.ahrq.gov/mepsweb/%20papers/st97/stat97.pdf>.

Potter, D.E.B. and Vistnes, J. (Editors). Health Insurance for the Elderly - Issues in Measurement Workshop, April 2003. Rockville (MD): Agency for Healthcare, Research, and Quality, 2006. MEPS Working Paper # 06003. Web only publication.

Rhoades, J.A.. Restricted-Activity Days in the United States, 1997 and 2001. Rockville (MD): Agency for Healthcare, Research, and Quality, 2004. Research Findings #22. AHRQ Pub. No. 04-0060.

Sam Zuvekas and Joel Cohen. A guide to comparing health care expenditures in the 1996 MEPS to the 1987 NMES. Inquiry. 2002 Spring;39(1):76-86.

Future Plans: Ongoing longitudinal data collection and dissemination will continue.

For more information:

Contact: MEPS Project Director

Phone: 301-427-1406

Email: [mepsprojectdirector@ahrq.hhs.gov](mailto:mepsprojectdirector@ahrq.hhs.gov)

Website: [www.meps.ahrq.gov/mepsweb](http://www.meps.ahrq.gov/mepsweb)

## **Centers for Disease Control and Prevention**

### **National Center for Chronic Disease Prevention and Health Promotion**

#### **Behavioral Risk Factor Surveillance System (BRFSS)**

Sponsor: Centers for Disease Control and Prevention (CDC)

Data Collection Agency/Organization: State-based systems

**Purpose:** The Behavioral Risk Factor Surveillance System (BRFSS) is designed to collect State-specific general population data on behaviors that are related to the leading causes of morbidity and mortality. States use BRFSS data to identify emerging health problems, establish and track health objectives, and develop and evaluate public health policies and programs. Because comparable methods are used from State to State and from year to year, States can compare risk factor prevalence with other States and monitor the effects of interventions over time. Also, the use of consistent methods in a large group of States permits the assessment of geographic patterns of risk factor prevalence. BRFSS data also can be used to examine smaller geographic areas within States. For example, CDC has analyzed BRFSS data starting in 2002 for many metropolitan and micropolitan statistical areas, and some counties within those areas for a project called Selected Metropolitan/Micropolitan Area Risk Trends (SMART) BRFSS. States and local health departments rely on data from the BRFSS to:

- Determine high-priority health issues and identify populations at highest risk for illness, disability, and death by analyzing data according to respondents' age, sex, education, income, and race/ethnicity.
- Develop strategic plan and targeted prevention activities and programs.
- Examine trends in behaviors over time to monitor the effectiveness of public health programs and progress in meeting prevention goals, such as Healthy People 2010 Objectives.
- Support community policies and programs that promote health and prevent disease—for example, by educating the public, the health community, and policymakers about disease prevention.

**Description:** The BRFSS is a State-based general adult population telephone surveillance system with data collected each month throughout the calendar year in all 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam with assistance from CDC. The BRFSS questionnaire is comprised of core questions and optional modules. There are three types of core questions. Fixed core questions are asked every year. Rotating core questions are asked every other year. Emerging core questions typically focus on “late-breaking” health issues. These questions are evaluated at the end of a survey year to determine if they are valuable. If the coordinators decide to keep the questions, they are added to the fixed core, rotating core, or optional modules, whichever is most appropriate. All states must ask all core questions. The optional modules are standardized questions that are supported by the CDC that cover additional health topics or are more detailed questions on a health topic included in the core. Each year states must choose which optional modules they will use based on the data needs of their state.



Current Activities: Data collection is ongoing.

Future Plans: States and urban areas will continue to rely on the BRFSS to gather the high quality data they need to plan and evaluate public health programs and to allocate scarce resources. CDC will work closely with State and Federal partners to ensure that the BRFSS continues to provide data that are useful for public health research and practice and for State and local health policy decisions.

Public Use Data Files: Data collection began in 1984 with 14 States. Data are available for each year 1984 through the present, but with a varied number of States participating until 1994. All data files and documentation are available on the BRFSS website at [http://www.cdc.gov/brfss/technical\\_infodata/index.htm](http://www.cdc.gov/brfss/technical_infodata/index.htm) under Technical Documents and Survey Data.

Associated Reports: Surveillance summaries and other relevant reports and information can be accessed at <http://www.cdc.gov/brfss/pubs/index.htm>.

For more information:

Email: [ccdinfo@cdc.gov](mailto:ccdinfo@cdc.gov)

Phone: 770-488-2455

Website: [www.cdc.gov/brfss](http://www.cdc.gov/brfss)

## **National Center for Health Statistics**

### **Longitudinal Studies of Aging (LSOAs)**

Sponsors: National Center for Health Statistics (NCHS), National Institute on Aging (NIA)

Data Collection Agency/Organization: U.S. Census Bureau and National Opinion Research Center at the University of Chicago.

Purpose: The LSOAs, a collaborative project of NCHS and NIA, is a family of surveys designed to measure changes in health status, health-related behaviors, health care, and the causes and consequences of these changes within and across two cohorts of elderly Americans. The project is comprised of four surveys: the 1984 Supplement on Aging, the 1984–1990 Longitudinal Study of Aging, the 1994 Second Supplement on Aging, and the 1994–2000 Second Longitudinal Study of Aging. An overview of each survey is provided below.

#### **Supplement on Aging (SOA)**

The SOA was conducted as part of the 1984 National Health Interview Survey (NHIS). The sample is comprised of 16,148 persons 55 years of age and over who were living in the community at the time of the interview. All interviews for this cross-sectional survey were conducted in person by the U.S. Census Bureau. The following topics were covered in the questionnaire:

- Housing characteristics
- Family structure and living arrangements
- Relationships and social contracts
- Use of community services
- Occupation and retirement (income sources)
- Health conditions and impairments
- Functional status, assistance with basic activities
- Utilization of health services, nursing home stays
- Health opinions

#### **Longitudinal Study of Aging (LSOA)**

The 1984 SOA served as the baseline for the LSOA, which followed all persons who were 70 years of age and over in 1984 (n=7,527) through three follow-up waves, conducted in 1986, 1988, and 1990. Follow-up interviews were conducted over the telephone by interviewers from the U.S. Census Bureau. The major focus of the follow-up interviews was on functional status and changes that had occurred between interviews. Information was also collected on housing and living arrangements, contact with children, utilization of health services and nursing home stays, health insurance coverage, and income.

## Second Supplement on Aging (SOA II)

The SOA II was conducted as part of the 1994 NHIS. Interviews were conducted during a follow-up visit with NHIS respondents between October 1994 and March 1996. The SOA II sample is comprised of 9,447 persons who had participated in the 1994 NHIS and had turned 70 years of age by the time of the SOA II interview. The SOA II serves as a comparison cohort to the 1984 SOA, and most of the questions from the SOA were repeated in the SOA II. Topics new to the SOA II include:

- Use of assistive devices and medical implants
- Health conditions and impairments
- Health behaviors
- Transportation
- Functional status, assistance with basic activities, unmet needs (expanded)
- Utilization of health services, nursing home stays (expanded)

## Second Longitudinal Study of Aging (LSOA II)

The SOA II serves as the baseline for the LSOA II. The LSOA II followed all 9,447 sample persons through two additional interviews, conducted in 1997–98 and 1999–2000. Follow-up interviews were conducted over the telephone by the National Opinion Research Center at the University of Chicago. The survey design and content of the LSOA II is similar to the first LSOA with improvements and enhancements reflecting the methodological and conceptual developments that have occurred during the intervening decade. The LSOA II data, when used in conjunction with data from the original LSOA, enables researchers to determine whether the prevalence and incidence of functioning, pathology, and impairments in the elderly population have changed over 10 years and whether the change is due to differences in cohort characteristics or to technological and medical advancements. Many of the questions included in the SOA II are repeated in the follow-up interviews. In addition, the follow-up questionnaires include items on cognitive functioning, income and assets, family and childhood health, and more extensive health insurance information.

The 1994–2000 LSOA II interview data have been linked with administrative data from NCHS' National Death Index and multiple cause of death files, with the Centers for Medicare & Medicaid Services, Medicare enrollment and claims data files, and with the Social Security Administration's benefit records.

### Data Availability:

#### SOA

Data from the full SOA sample of persons 55 years of age and older are available on CD-ROM: Series 10, No. 16H (issued July 2001). This CD-ROM also includes data obtained during the 1984 NHIS interview and all related documentation. The study is fully documented in Vital and Health Statistics, Series 1 Number 18.

## LSOA

The complete set of data from all four interview waves and corresponding documentation is available on CD-ROM: The Longitudinal Study of Aging, 1984–1990, No. 1 (issued September 1993). The study is fully documented in three Vital and Health Statistics monographs: Series 1, Numbers 18, 21, and 28.

## SOA II

The SOA II is available on CD-ROM: The Second Supplement on Aging, 1994, Version 2, No. 1 (issued September 1998).

## LSOA II

The complete set of LSOA II data is available online from the LSOAs website. Data users interested in these files are directed to <http://www.cdc.gov/nchs/lsoa.htm>.

To request copies of the LSOA CD-ROMs mentioned above, email [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov) or telephone (301) 458-INFO. Members of ICPSR at the University of Michigan may obtain the data free of charge through ICPSR.

**Linked Data:** Information about the LSOA II data linkage projects mentioned above and availability of resulting data files may be obtained online from the NCHS Data Linkage Activities website. Interested users are directed to [http://www.cdc.gov/nchs/data\\_access/data\\_linkage\\_activities.htm](http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm).

**Future Plans:** While active data collection and passive followup of the LSOA II cohort is complete, documentation and research are ongoing. Users will find updated items at the LSOAs website (<http://www.cdc.gov/nchs/lsoa.htm>).

**Reports:** Detailed descriptions of the LSOA survey are published in reports of the Vital and Health Statistics, Series 1, Number 21 and Number 28. The LSOA questionnaires are provided in Appendices III-V of Series 1, Number 28. All LSOA II documentation is available online at <http://www.cdc.gov/nchs/lsoa.htm>. Documentation associated with the linked data files is available at [http://www.cdc.gov/nchs/data\\_access/data\\_linkage\\_activities.htm](http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm).

For more information:

Contact: Julie Dawson Weeks

Email: [lsoa@cdc.gov](mailto:lsoa@cdc.gov)

Phone: 301-458-4562

Website: [www.cdc.gov/nchs/lsoa.htm](http://www.cdc.gov/nchs/lsoa.htm)

## **National Ambulatory Medical Care Survey (NAMCS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: U.S. Census Bureau

**Purpose:** The National Ambulatory Medical Care Survey (NAMCS) collects data on the utilization and provision of medical care services at visits to office-based physicians. Data are collected on type of providers seen; reason for visit; diagnoses; drugs ordered, provided, or continued; diagnostic and screening services, and therapeutic services ordered or provided during the visit, as well as other visit characteristics. Patient data include age, sex, race, and expected source of payment. Data are also collected on selected characteristics of physicians included in the survey.

**Survey Universe:** The survey is a representative sample of visits to nonfederally employed office-based physicians, not including anesthesiologists, radiologists, or pathologists, who are primarily engaged in direct patient care. Telephone contacts are excluded.

**Research Design:** Census field representatives contact sample physicians to determine if they are eligible for survey participation. Visits to eligible physicians are systematically sampled over a 1-week reporting period such that about 30 encounters are selected.

**Survey Mode:** Physicians are asked to complete Patient Record Forms (PRF) for each sampled visit, but Census field representatives typically abstract data for about thirty percent of these visits.

**Unit of Analysis:** Visit-level and physician-level.

**Sample:** In any given year, the initial sample consists of approximately 3,000 physicians stratified into 15 specialty groups. In 2003, 67 percent of eligible physicians responded, and 25,288 PRFs were submitted. In 2004, 65 percent of eligible physicians responded, and 25,286 PRFs were submitted.

**Topics:** Medical care, physician visits, diagnoses, prescription drugs.

**Data Availability:** Annual data collection began in 1973 and continued through 1981. The survey was next conducted in 1985, and resumed an annual schedule in 1989. National and regional (four Census regions) estimates are available.

**Linked Data:** Restricted NAMCS data contain geographical variables such as patient ZIP code and State and county codes. These variables can be linked to other data sources such as U.S. Census Bureau data and the Health Resources and Services Administration's Area Resource File, a national county-level health resources information system, in order to do contextual analysis. Such linking can only be done at NCHS' Research Data Center.

Data Dissemination: Public-use data files are available for 1973–2006. In addition, SAS program code is available for survey years 1993–2006; SPSS and Stata code are available for 2002–0. Data for 1990–2003 are also available on CD-ROM with Statistical Export and Tabulation System (SETS) retrieval software.

Reports: A list of publications and reports using NAMCS data is available at:  
<http://www.cdc.gov/nchs/namcs.htm>.

Future Plans: The NAMCS PRF is modified approximately every 2 years to reflect changes in physician practice characteristics, patterns of care, and technological innovations. Examples of recent changes are the number of drugs recorded on the PRF, and checkboxes of specific tests or procedures performed. Added for the first time in 2006 were a sample of community health centers, a sample of oncologists, and a cervical cancer screening supplement.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: [www.cdc.gov/nchs/namcs.htm](http://www.cdc.gov/nchs/namcs.htm)

## **National Death Index (NDI)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: Division of Vital Statistics, NCHS

**Purpose:** The National Death Index (NDI) is a central computerized index of death record information on file in the State vital statistics offices. Working with these State offices, NCHS established the NDI as a resource to aid epidemiologists and other health and medical investigators with their mortality ascertainment activities. Investigators conducting prospective and retrospective studies can use the NDI to determine whether persons in their studies may have died and to obtain the underlying and multiple causes of death in coded form (for the more probable matches).

It is available to investigators solely for statistical purposes in medical and health research. Not accessible to organizations or the general public for legal, administrative, or genealogy purposes.

The NDI currently contains death records from 1979–2006. Death records are added to the NDI file annually, approximately 12–15 months after the end of a particular calendar year. Deaths for 2007 should be available in May 2009.

Refer to the NDI matching criteria ([http://www.cdc.gov/nchs/data/NDICriteria\\_Front.pdf](http://www.cdc.gov/nchs/data/NDICriteria_Front.pdf)) to see how records are selected as possible matches. NDI users are encouraged to submit as many of the following data items as possible for each study subject: first and last name, middle initial, father's surname, social security number, month, day, and year of birth, race, sex, marital status, State of residence, and State of birth.

Refer to the NDI retrieval report ([http://www.cdc.gov/nchs/data/NDI\\_Retrieval\\_Back.pdf](http://www.cdc.gov/nchs/data/NDI_Retrieval_Back.pdf)) for a sample of how possible matches are presented.

Investigators can then make arrangements with the appropriate State offices to obtain copies of death certificates or specific statistical information such as cause of death. Investigators can also obtain cause of death codes directly using the NDI Plus service.

To use the system, investigators first must submit a NDI application form to NCHS. Applicants should allow about 2 months for their applications to be reviewed and approved. Once approved, users may submit their study subjects' names, social security numbers, dates of birth, and related information to NCHS on diskette or CD-ROM.

For more information:

Email: [ndi@cdc.gov](mailto:ndi@cdc.gov)

Phone: 301-458-4444

Website: [www.cdc.gov/nchs/ndi.htm](http://www.cdc.gov/nchs/ndi.htm)

## **National Health and Nutrition Examination Survey (NHANES)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: Westat

Purpose: NHANES is designed to assess the health and nutritional status of the U.S. noninstitutionalized civilian population through direct physical examinations, laboratory tests, and interviews. NHANES collects data on the prevalence of selected disease and risk factors, as well as a range of other topics.

Survey Universe: NHANES I included the civilian noninstitutionalized population 1 year of age and over residing in the contiguous United States, except for people residing on any of the reservation lands set aside for the American Indians. NHANES II targeted the civilian noninstitutionalized population 6 months of age and over residing in the United States, including Alaska and Hawaii. The Hispanic Health and Nutrition Examination Survey (HHANES) included three geographically and ethnically distinct populations: 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. NHES I, NHANES I, and NHANES II collected information on persons up to 74 years of age. NHANES III and later surveys include people age 75 years and over. In addition, NHANES III provides estimates for the civilian noninstitutionalized population 2 months of age and over in the United States.

Research Design: This is a cross-sectional survey first conducted in 1960, and has been conducted continuously since 1999. In 1971, the nutritional surveillance component was added, and the survey changed from NHES (National Health Examination Survey) to NHANES.

Survey Mode: The NHANES includes clinical examinations and selected medical and laboratory tests that are conducted in mobile examination centers (MECs). Prior to 2003, a small number of participants who were unable to come to the MEC received an abbreviated health examination in their homes. In addition, in-home personal interviews are conducted.

Unit of Analysis: Person-level data analysis.

Sample: The survey for the NHANES III was conducted from 1988 to 1994 and consisted of two phases of equal length and sample size. Phase 1 and Phase 2 comprised random samples of the civilian U.S. population living in households. About 40,000 persons 2 months of age and over were selected and asked to complete an extensive interview and an examination. Participants were selected from households in 81 counties across the United States. Children aged 2 months to 5 years and persons 60 years of age and over were oversampled to provide precise descriptive information on the health status of selected population groups of the United States. Beginning in 1999 NHANES oversampled low-income persons, adolescents 12–19 years of age, 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.

Over the 6-year survey period of NHANES III, 39,695 persons were selected, the household interview response rate was 86 percent, and the medical examination response rate was 78 percent. In the sample selection for NHANES 1999–2000, there were 22,839 dwelling units screened. Of these, 6,005 households had at least one eligible sample person identified for interviewing. A total of 12,160 eligible sample persons were identified. The overall response rate in NHANES 1999–2000 for those interviewed was 81.9 percent (9,965 of 12,160), and the response rate for those examined was 76.3 percent (9,282 of 12,160). For NHANES 2001–02 there were 13,156 persons selected in the sample, of which 83.9 percent (11,039) were interviewed and 79.7 percent (10,480) of the 13,156 selected completed the health examination component of the survey.

Topics: Cardiovascular disease, diabetes, environmental exposures, hearing loss, infectious disease, kidney disease, mental health, cognitive functioning, nutrition, obesity, oral health, osteoporosis, physical fitness, physical functioning, reproductive history, sexual behavior, respiratory disease, sexually transmitted diseases, vision.

Data Availability: Data have been collected from surveys conducted during 1960–62 (NHES I), 1963–65 (NHES II), 1966–70 (NHES III), 1971–74 (NHANES I), 1976–80 (NHANES II), 1982–84 (NHANES), and 1988–94 (NHANES III). Beginning in 1999, the survey has been conducted continuously. National-level data are available and are released in 2 year cycles (1999-2000; 2001-2002; 2003-2004 etc.).

Linked Data: NCHS has conducted a linkage of NHANES I Epidemiologic Followup Study (NHEFS) with the NCHS National Death Index (NDI). The NHEFS is a longitudinal follow-up study of the adult participants from the first National Health and Nutrition Examination Survey (NHANES I) and includes all participants 25–74 years of age who completed a medical examination during the NHANES I survey period between 1971 and 1975. A linkage is available of the Second National Health and Nutrition Examination Survey (NHANES II) with the NCHS National Death Index (NDI). The NHANES II mortality linkage includes information on 9,252 participants 30–75 years of age who completed a medical examination during the NHANES II survey period between 1976 and 1980. Finally, there is a mortality linkage of the Third National Health and Nutrition Examination Survey (NHANES III) with the NDI. The NHANES III linked mortality file provides mortality follow-up data from the date of NHANES III survey participation (1988–94) through December 31, 2000. All NHANES III participants are included on the linked mortality files but only adult participants (those 17 years and over) were eligible for mortality follow-up.

Data Disseminated: Data from NHANES 1999+, NHANES III, NHANES II, and NHEFS are available online via the NCHS website.

Reports:

- Surveillance for Dental Caries, Dental Sealants, Tooth Retention, Edentulism, and Enamel Fluorosis—United States, 1988–94 and 1999–2002
- Lower Extremity Disease Among Persons Aged greater than/equal to 40 Years With and Without Diabetes—United States, 1999–2002
- Mobility Limitation Among Persons Aged greater than/equal to 40 Years With and Without Diagnosed Diabetes and Lower Extremity Disease—United States, 1999–2002
- Blood Lead Levels—United States, 1999–2002
- Children and Teens Told by Doctors That They Were Overweight—United States, 1999–2002
- DHHS-USDA Dietary Survey Integration—What We Eat In America
- National Center for Environmental Health’s Third National Report on Human Exposure to Environmental Chemicals
- Healthy Eating Index, USDA Center for Nutrition Policy and Promotion

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: <http://www.cdc.gov/nchs/nhanes.htm>

## **National Health and Nutrition Examination Survey—Epidemiologic Followup Study (NHEFS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: Westat

Purpose: NHEFS was designed to investigate the relationships between clinical, nutritional, and behavioral factors assessed in the first National Health and Nutrition Examination Survey (NHANES I) and subsequent morbidity, mortality, and hospital utilization, as well as, changes in risk factors, functional limitation and institutionalization.

Research Design: The NHEFS cohort includes all persons between 25 and 74 years of age who completed a medical examination at NHANES I in 1971–75 (n=14,407). It is comprised of a series of followup studies, four of which have been conducted to date. The first wave of data collection was conducted for all members of the NHEFS cohort from 1982–84. It included tracing the cohort; conducting personal interviews with subjects or their proxies; measuring pulse rate, weight and blood pressure of surviving participants; collecting hospital and nursing home records of overnight stays; and collecting death certificates of decedents.

Continued followups of the NHEFS population were conducted in 1986, 1987, and 1992 using the same design and data collection procedures developed in the 1982–84 NHEFS, with the exception that a 30-minute computer-assisted telephone interview was administered rather than a personal interview, and no physical measurements were taken. The 1986 NHEFS, was conducted for members of the NHEFS cohort who were 55–74 years of age at their baseline examination and not known to be deceased at the 1982–84 NHEFS (n=3,980). The 1987 NHEFS was conducted for the entire nondeceased NHEFS cohort (n=11,750). The fourth wave of data collection, the 1992 NHEFS, was conducted for the entire nondeceased NHEFS cohort (n=11,195).

Current Activities: NCHS has collected mortality data through 2000, Medicare entitlement and utilization data (1991–2000) and Social Security benefit history data (1962–2003) for eligible NHEFS survey participants. For more information visit [http://www.cdc.gov/nchs/data\\_access/data\\_linkage\\_activities.htm](http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm).

Future Plans: While no full scale interview recontacts are currently planned for this cohort, NCHS plans to continue to link NHEFS survey data to administrative records on mortality and health care utilization.

Data Dissemination: Data from the NHEFS can be downloaded from the NCHS website <http://www.cdc.gov/nchs/about/major/nhefs/nhefs.htm>.

Reports: Descriptions of the study methodologies are available in the NCHS Vital and Health Statistics Series 1 Reports, Numbers 22, 25, 27, and 35. Statistical issues in analyzing NHEFS data are addressed in Series 2 Report, Number 121. See <http://www.cdc.gov/nchs/about/major/nhefs/nhefs.htm>.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: [www.cdc.gov/nchs/about/major/nhefs/nhefs.htm](http://www.cdc.gov/nchs/about/major/nhefs/nhefs.htm)

## National Health Interview Survey (NHIS)

Sponsor: National Center for Health Statistics; special topic questions are sponsored by various agencies

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: The NHIS is the principal source of information on the health of the population of the United States. The NHIS monitors the health of the U.S. population through the collection and analysis of data on a broad range of data and is used to monitor trends in illness and disability and to track progress toward achieving national health objectives. The data are also used by the public health research community for epidemiologic and policy analysis of such timely issues as characterizing those with various health problems, determining barriers to accessing and using appropriate health care, and evaluating Federal health programs.

Survey Universe: This survey includes the civilian noninstitutionalized population, excluding patients in long-term care facilities, persons on active duty with the Armed Forces, prisoners, and U.S. nationals living in foreign countries. The institutionalized population of the United States accounts for approximately 2 percent of the total population.

Research Design: The NHIS has been conducted annually since 1957, with a major redesign every 10–15 years. It is a cross-sectional household interview survey; sampling and interviewing are continuous throughout the year. The last major revision to the instrument occurred in 1997 and it divided the survey into three basic parts. The Basic module consists of the family core, sample adult core, and sample child core. The other two modules are the Periodic module and the Topical module.

Survey Mode: The data are collected through personal household interview surveys conducted by trained interviewers. Prior to 1997, a paper and pencil questionnaire format was used. From 1997 onwards, computer-assisted personal interviewing (CAPI) was used. The last re-engineering of the CAPI occurred in 2004.

Unit of Analysis: Person-level analysis: Self report and proxy report.

Sample: A new sample design for the NHIS was implemented in 2006 and will be in effect for several years. The fundamental structure of the new 2006 NHIS sample design is very similar to the previous 1995-2005 NHIS sample design. State-level stratification is retained in the new NHIS sample design, which allows use of the NHIS for producing state estimates and for possible *future* dual-frame surveys at the state level. (Users should note that the current NHIS sample size is not sufficient to provide reliable state-level estimates for most states; however, for those states with larger populations, reliable state-level estimates can be produced. Since state identifiers are not publicly released, use of that information can be made through the NCHS Research Data Center. Contact the NCHS Research Data Center for more information, or visit their Web page: <http://www.cdc.gov/nchs/r&d/rdc.htm>). The new sample design reduced the NHIS sample size by about 13%.

Oversampling of the black and Hispanic populations was retained in the 2006 redesign to allow for more precise estimation of health characteristics in these growing minority populations. The new sample design also oversamples the Asian population. In addition, the sample adult selection process has been revised so that when black, Hispanic, or Asian persons aged 65 years or older are present, they have an increased chance of being selected as the sample adult.

Topics: The Core NHIS includes topics on injuries, illnesses, chronic conditions, risk factors, health behaviors, health status, health insurance coverage, health care utilization, demographics, and socioeconomic status. Recent supplements have included topics on Complementary and Alternative Medicine, Cancer Prevention, and Healthy People 2010 objectives.

Data Availability: National-level data are available annually since 1957. The NHIS sample is not designed to provide State level data with acceptable precision for every State.

Linked Data (available through the NCHS Research Data Center):

- NHIS files are linked to the Medical Expenditure Panel Survey for the years 1996–2002 (MEPS) and 1995–2001 (NHIS)
- NHIS 1986–2000 are linked to mortality files from 1986 to 2002
- In 1994, the NHIS data was linked with National Immunization Provider Record Check Study to determine the accuracy of household vaccination reports for children 12–35 months of age
- NHIS has been linked with 1962–2003 Social Security administrative data
- NHIS has been linked with 1991–2000 Medicare administrative data

Data Dissemination: Microdata files are available via the Internet for 1963–2007. Questionnaires for 2008 and 2009 are also available on the Internet.

Reports: An NHIS bibliography via MEDLINE, Advanced data Reports, and Vital and Health Statistics, Series 10 Reports are available at: <http://www.cdc.gov/nchs/nhis.htm#Publications>.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: [www.cdc.gov/nchs/nhis.htm](http://www.cdc.gov/nchs/nhis.htm)

## **National Home and Hospice Care Survey (NHHCS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: Westat

**Purpose:** The NHHCS collects data on the characteristics and care provided by home health care agencies and hospices as well as characteristics of patients receiving these services. Data collected at the home health care and hospice agency level include number of clients served, ownership and affiliations, certification status, and services provided. At the patient level, data are collected on demographic characteristics, diagnoses, living arrangements, caregiver status, enrollment date, discharge disposition (for discharge sample), selected therapies and treatments provided, aids and special devices used, activities of daily living (ADL) assistance received from the agency, vision and hearing impairments, continence, payment source, and care charges.

**Survey Universe:** The survey covers agencies (licensed or certified by Medicare or Medicaid) and the current patients and discharges from agencies that provide home health and hospice care services in the United States. Agencies may be freestanding health facilities or units of larger organizations, such as hospitals or nursing homes. Agencies providing only durable medical equipment are excluded.

**Research Design:** This is a continuing series of cross-sectional surveys.

**Survey Mode:** Agency characteristics were obtained through interviews with the agency administrators and staff. Sample patients and discharges were selected, and questionnaires were completed by interviewing the staff member most familiar with the care provided to the patient. Respondents were requested to refer to the medical records for the patient.

**Unit of Analysis:** Home health care and hospice care agency level and patient level.

**Sample:** 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 up to six current patients and six 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 12 months before the survey.

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 from a universe of home health agencies and hospices obtained from various national organizations and other sources. In 2000, 1,800 agencies were sampled from the universe that was obtained from SMG Home Healthcare Market Database and the membership list of the National Hospice and Palliative Care Organization. The response rates during the 1992–2000 survey years have been greater than 92 percent at the agency level, mid-90 to mid-80 percent for current patients, and low 90 to low 80 percent for discharges.

Topics: Home health care, hospice, disability.

Data Availability: Initiated in 1992, the NHHCS was also conducted in 1993, 1994, 1996, 1998, and 2000. The survey was conducted again in 2007. Data release of the 2007 survey is expected in spring 2009.

Linked Data: In 2007, NHHCS data from sampled patients and agencies will be linked to the CMS Outcome Assessment Information Set (OASIS) to obtain more clinical data and create episodes of care.

Data Dissemination: CD-ROMs are available from National Technical Information Service (NTIS) for 1992–94, 1996, 1998 and 2000. Data from 1996-2000 can be downloaded from the NHHCS webpage.

Reports: A list of reports is available at: <http://www.cdc.gov/nchs/products.htm#reports>.

Future Plans: Major design and content changes are scheduled for the 2007 NHHCS. A new data collection system will be employed: Computer-assisted Personal Interview (CAPI) System. In addition, new content will be added, including agency characteristics such as agency accreditation, electronic information systems, cultural competency, immunization policies and practices, end-of-life practice, and education, certification, and tenure of agency administrators, directors of nursing, and medical directors. New patient characteristics such as pain assessment and pain relief and new staffing information such as specialty certification, nursing turnover and stability, recruitment and retention strategies, and entry level wage will also be added. There will be a new supplement, National Home Health Aide Survey, which will collect information from a sample of participating agencies and will collect information on job satisfaction, organizational culture, recruitment and training, family life and injuries. This data will be collected via Computer-assisted Telephone Interview (CATI).

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: <http://www.cdc.gov/nchs/nhhcs.htm>



## **National Hospital Ambulatory Medical Care Survey (NHAMCS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: U.S. Census Bureau

**Purpose:** The National Hospital Ambulatory Medical Care Survey (NHAMCS) collects data on the utilization and provision of medical care services in hospital emergency and outpatient departments. Data are collected from medical records on type of providers seen; reason for visit; diagnoses; drugs ordered, provided, or continued; diagnostic and screening services, selected procedures and tests ordered or performed during the visit, and other visit characteristics. Patient data include age, sex, race, and expected source of payment. Data are also collected on selected characteristics of hospitals included in the survey.

**Survey Universe:** The survey is a representative sample of visits to emergency departments (EDs) and outpatient departments (OPDs) of non-Federal, short stay, and general hospitals. Telephone contacts are excluded.

**Research Design:** Census field representatives contact sample hospitals to determine whether they have a 24-hour ED or an OPD that offers physician services. Visits to eligible EDs and OPDs are systematically sampled over the 4-week reporting period such that about 100 ED encounters and about 200 OPD encounters are selected.

**Survey Mode:** Hospital staff is asked to complete a Patient Record Form (PRF) for each sampled visit, but Census field representatives typically abstract data for more than half of these visits.

**Unit of Analysis:** Visit-level and hospital-level.

**Sample:** In any given year, the hospital sample consists of approximately 500 hospitals, of which 80 percent have EDs and about half have eligible OPDs. Typically, about 1,000 clinics are selected from participating hospital OPDs.

**Topics:** Hospitals, medical care, diagnoses.

**Data Availability:** Annual data collection began in 1992. National and regional (four Census regions) estimates are available.

**Linked Data:** Restricted NHAMCS data contain geographical variables such as patient ZIP code and State and county codes. These variables can be linked to other data sources such as U.S. Census Bureau data and the Health Resources and Services Administration's Area Resource File, a national county-level health resources information system, in order to do contextual analysis. Such linking can only be done at NCHS' Research Data Center.

**Data Dissemination:** Public-use data files are available for 1992-2006.

Reports: A list of publications and reports using NHAMCS data is available at:  
<http://www.cdc.gov/nchs/nhamcs.htm>.

Future Plans: The NHAMCS PRF is modified approximately every 2 years to reflect changes in physician practice characteristics, patterns of care, and technological innovations. Examples of recent changes are the number of drugs recorded on the PRF form, and checkboxes of specific tests or procedures performed. In addition, supplemental studies are conducted periodically to assess topics such as bioterrorism preparedness, hospital staffing and capacity, and ambulance diversions. For 2006, a supplement on emergency pediatric surveillance and equipment and an outpatient cervical cancer screening supplement are being fielded. Data from the supplements are currently available only through NCHS' Research Data Center.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: [www.cdc.gov/nchs/nhamcs.htm](http://www.cdc.gov/nchs/nhamcs.htm)

## **National Hospital Discharge Survey (NHDS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: NCHS, sample of inpatient hospital discharges, U.S. Census Bureau.

Purpose: The National Hospital Discharge Survey (NHDS) collects and produces national estimates on characteristics of inpatient stays in non-Federal short-stay hospitals in the United States. Patient information collected includes demographics, length of stay, diagnoses, and procedures. Hospital characteristics collected include region, ownership, and bed size.

Survey Universe: The survey design covers the 50 States and the District of Columbia. Included in the survey are hospitals with an average length of stay of less than 30 days for all inpatients, general hospitals, and children's general hospitals. Excluded are 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. All discharged patients from in-scope hospitals are included in the universe from which the sample is drawn.

Research Design: Beginning in 1985, two data collection procedures have been used in the survey. One is a manual system in which sample selection and medical transcription from the hospital records to abstract forms is performed by the hospital's staff or by staff of the U.S. Bureau of the Census on behalf of NCHS. The other data collection procedure is an automated system in which NCHS purchases machine-readable medical record data from commercial organizations, State data systems, hospitals, or hospital associations. The medical abstract form and the automated data tapes contain items that relate to the personal characteristics of the patient. These items include age, sex, race, ethnicity, marital status, and expected sources of payment. Administrative items such as admission and discharge dates (which allow calculation of length of stay), as well as discharge status, are also included. Medical information about patients includes diagnoses and procedures coded to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM).

A redesign of the NHDS was implemented for the 1988 survey. Under the redesign hospitals were selected using a modified three-stage stratified design.

Survey Mode: From 1965, the initial year of the NHDS, through 1984, all data collection was conducted by means of manual abstraction of patient information from sampled medical records. Sample selection and transcription of information from inpatient medical records to NHDS survey forms were performed by either hospital staff or staff of the Census Bureau on behalf of NCHS. When the second data collection procedure was introduced in 1985, that is, using automated data, the new method was used in approximately 17 percent of the sample hospitals for 1985–87. Discharges from these data files are selected using the NHDS sampling specifications, and the same computer edits and estimation procedures used for the manual data are also used for the automated data. Two data collection methods, manual and automated,

continue to be used in the NHDS. For the 2004 data year, approximately 44 percent of respondent hospitals provided data through the automated system.

**Unit of Analysis:** The basic unit of estimation for NHDS is a sampled discharge.

**Sample:** The NHDS hospital sample is updated every 3 years by continuing the sampling process among hospitals that become eligible for the survey during the intervening years and by deleting hospitals that were no longer eligible. This process has been conducted every 3 years beginning in 1970, and will be conducted again in 2006.

In 2004, 501 hospitals were selected: 476 were in scope, 439 participated (92 percent), and approximately 371,000 medical records were abstracted.

**Topics:** Data from the NHDS are used for evaluating the health status of the population, planning programs to improve health status, studying trends in morbidity, and carrying out other research in epidemiology, public health, and health services research. Specific topics recently explored include asthma, injury, obstetric procedures, and heart disease.

**Data Availability:** The NHDS has been conducted annually since 1965. National and regional (four Census regions) estimates, based on the calendar year, are produced.

**Linked Data:** Linkages of NHDS data to other data files is possible through the National Center for Health Statistics' Research Data Center. Among the data files, which have been linked to the NHDS are area resource file, American Hospital Association (AHA), and U.S. Census Bureau data.

**Data Dissemination:** Public-use data files are available for download from the website for 1996–2006. NHDS data is also available on CD-ROM's for data years 1970–97. Reports: Reports utilizing data from NHDS can be found at:  
<http://www.cdc.gov/nchs/about/major/hdasd/listpubs.htm>.

**Future Plans:** A contract for developing a conceptual framework for the redesign of the NHDS and for developing the methods and procedures for carrying it out is underway. In 2007 a feasibility study of the methods will be undertaken in nine hospitals. Following the revision of the methods based upon the feasibility study, a new contract will be awarded in 2008 for a pilot test of hospitals in a number of States.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 301-458-4321

Website: [www.cdc.gov/nchs/about/major/hdasd/nhds.htm](http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm)

## **National Mortality Followback Survey (NMFS)**

Sponsor: National Center for Health Statistics. Supplemental questions are sponsored by various agencies.

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To provide information on mortality beyond that routinely collected on the death certificate. Five major subject areas can be examined: (1) socioeconomic differentials in mortality; (2) prevention of premature death by inquiring into the association between risk factors and the cause of death; (3) health care service utilization in the last year of life; (4) disability and medical conditions in the last year of life; and (5) the reliability and validity of certain items reported on the death certificate.

The Mortality Followback Survey Program, begun in the 1960s by NCHS, uses a sample of United States residents who die in a given year to supplement the death certificate with information from the next of kin or another person familiar with the decedent's life history. This information, sometimes enhanced by administrative records, provides a unique opportunity to study the etiology of disease, substance use, demographic characteristics and risk factors in mortality, and other health related issues.

The first mortality followback survey, conducted in 1961, featured information on hospital and institutional care in the last year of life. Information from the 1962–63 survey permitted an extensive analysis of socioeconomic differentials in mortality. Data from the 1964–65 survey included expenditures for health care during the last year of life, sources of payment, and health insurance coverage of decedents. The 1966–68 survey provided information on the link between smoking and cancer mortality. In 1986, the survey provided data on comorbid conditions, disabilities, alcohol use, and access to health care services.

### **The 1993 National Mortality Followback Survey**

The 1993 survey samples individuals aged 15 years or over who died in 1993. Forty-nine of the 50 State vital registration areas granted approval to sample their death certificates, as well as the independent vital registration areas of the District of Columbia and New York City. (South Dakota declined to participate in the NMFS due to State law restricting the use of death certificate information.)

A sample of 22,957 death certificates from the Current Mortality Sample was drawn. To meet specific research needs, the sample included 9,636 death certificates selected with certainty. There is an over-sample of death certificates to obtain reliable numbers for important population subgroups; persons under age 35, women, and the black population.

The 1993 NMFS focused on five subject areas:

- Socioeconomic differentials in mortality
- Associations between risk factors and cause of death

- Disability
- Access and utilization of health care facilities in the last year of life
- Reliability and validity of certain items reported on the death certificate
- Identifying strategies to prevent deaths due to trauma

The 1993 NMFS is different from the five previous mortality followback surveys in several ways:

- It over sampled deaths due to homicide, suicide, and unintentional injury.
- The subject areas are considerably broader. However, many previously-surveyed subject areas are included for prevalence analysis from various years.
- The survey is the first national survey to collect information from medical examiner/coroner reports.
- The complexity of the questionnaire necessitated telephone or in person interviews.

The 1993 NMFS was designed in collaboration with other agencies of the Public Health Service, Department of Health and Human Services, and the National Highway Traffic Safety Administration. Several of these agencies provided funding through NCHS's Reimbursable Work Program. Results from the first release of data from the 1993 NMFS are available on the FTP server.

Data Dissemination: The NMFS can only be purchased through the National Technical Information Service (NTIS). Contact NTIS for current pricing and availability. See [http://www.cdc.gov/nchs/products/elec\\_prods/subject/nmfs.htm](http://www.cdc.gov/nchs/products/elec_prods/subject/nmfs.htm).

Reports: Publications using the 1986 NMFS include: Advance Data Nos. 172, 173, and 180. Vital and Health Statistics Series 1, No. 29; Series 2, No. 118; and Series 20, No. 19.

For more information:

Contact: Mortality Statistics Branch

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 301-458-4666

Website: [www.cdc.gov/nchs/about/major/nmfs/nmfs.htm](http://www.cdc.gov/nchs/about/major/nmfs/nmfs.htm)

## **National Nursing Home Survey (NNHS)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: Westat

**Purpose:** The NNHS provides information on characteristics of nursing homes and their residents and staff. The NNHS provides information on nursing homes from two perspectives: that of the provider of services and that of the recipient. Data about the facilities include characteristics such as bed size, ownership, affiliation, Medicare/Medicaid certification, specialty units, services offered, number and characteristics of staff, expenses, and charges. Data about the current residents and discharges include demographic characteristics, health status, level of assistance needed with activities of daily living, vision and hearing impairment, continence, services received, sources of payment, and discharge disposition (for discharges). The survey underwent a major redesign in 2004. New content added to the survey included medications, medical, mental health, and dental services offered or provided, end-of-life care and advance directives, education, specialty credentials, and length of service of key staff, turnover and stability of nursing staff, use of contract/agency staff, overtime shifts worked, wages and benefits, facility practices for immunization, dining, and use of mechanical lifting devices.

**Survey Universe:** The initial NNHS, conducted in 1973–74, included the universe of nursing homes that provided some level of nursing care and excluded homes providing only personal or domiciliary care. The 1977 NNHS encompassed all types of nursing homes, including personal care and domiciliary care homes. The 1985 NNHS was designed to be similar to the 1973–74 survey in that it excluded personal or domiciliary care homes. The 1995, 1997, 1999, 2004 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.

Data were collected from nursing homes in all 50 States and the District of Columbia in the 1995–2004 surveys, but in 1973–74, 1977, and 1985, data were only collected in the 48 contiguous States and DC. Data on current residents were collected in all surveys; data on discharges were collected in 1977, 1985, 1997, and 1999.

**Research Design:** NNHS is a continuing series of cross-sectional surveys. The 2004 NNHS used a self-administered questionnaire to obtain staffing information and a computer-assisted personal interview system (CAPI), which contained seven modules: facility qualification, facility characteristics, resident sampling, nursing assistant sampling, health assessment 1, health assessment 2, prescription medications, and payment.

**Survey Mode:** Information on the facility is collected through a personal interview with the administrator or staff designated by the administrator. Resident data were provided by staff familiar with the care provided to the resident. Staff 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, in several years staffing data were collected via a self-administered

questionnaire. Discharge data were based on information recorded in the medical record. Information on discharges was not collected in 1995 and 2004.

Unit of Analysis: Information is collected on the nursing home level and resident level.

Sample: The 2004 sample consisted of 1,496 nursing homes. In 1995, 1997, and 1999, facility-level response rates were over 93 percent.

Topics: Nursing homes, health status, disability, prescription drugs.

Data Availability: NCHS conducted six NNHS: the first survey August 1973–April 1974; the second May–December 1977; the third August 1985–January 1986; the fourth July–December 1995; the fifth July–December 1997; and the sixth July–December 1999. The most recent NNHS was conducted in 2004 and data were released in August 2006.

Linked Data: The 2004 NNHS data from sampled residents and facilities will be linked to the CMS Minimum Data Set (MDS) to obtain more clinical data and create episodes of care. The NNHS data are also linked to the NDI.

Data Dissemination: CD-ROMs are available from National Technical Information Service (NTIS) for 1999, 1997, 1995, 1985, 1977, 1973–74, and 1969. Data for 1995–2004 can be downloaded from the NNHS Web page.

Reports: A list of reports is available at: <http://www.cdc.gov/nchs/products.htm#reports>.

Future Plans: The 2004 NNHS has been redesigned and expanded to better meet the data needs of researchers and health care planners working to ensure that quality long-term care will be available for the nation's growing senior population. The survey will utilize computer-assisted personal interviewing. This computerized system speeds the flow of data making it possible to release information on a timelier basis and makes it easier for respondents to participate in the survey. Another change is an increase in the current resident sample size, allowing more detailed and better information to be collected about this population.

The survey has been expanded to include new and critical topics, such as:

- Medications and adverse experiences with medications
- Medical, mental health and dental services provided residents
- End-of-life care and advance directives
- Education, specialty credentials, and length of service for key staff
- Turnover and stability of nursing staff; use of temporary/agency staff; overtime shifts worked, and benefits
- Facility practices for immunizations, dining, and use of mechanical lifting devices

The Nursing Assistant Supplement to the National Nursing Home Survey will provide information needed to recruit, retain, and develop this essential paraprofessional long-term care workforce. The survey is designed to determine the likelihood that workers will continue in their present positions and the factors that affect those decisions, including job satisfaction,



environment, training, advancement opportunities, benefits, working conditions, and personal or family demands. This first national survey of nursing assistants will be conducted as a separate telephone interview with a sample of workers who provide nursing home residents assistance with Activities of Daily Living (eating, transferring, toileting, dressing, and bathing).

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: <http://www.cdc.gov/nchs/mnhs.htm>

## **National Nursing Home Survey Follow-up (NNHSF)**

Sponsors: National Center for Health Statistics, National Institute on Aging

Data Collection Agency/Organization: Research Triangle Institute

Purpose: The primary purpose of the NNHSF is to provide data on the flow of persons in and out of long-term care facilities and hospitals.

Research Design: The National Nursing Home Survey Follow-up (NNHSF) is a longitudinal study that follows the cohort of current residents and discharged residents sampled from the 1985 National Nursing Home Survey (NNHS). The NNHSF builds on the data collected in the 1985 NNHS by providing longitudinal information on nursing home and hospital utilization. The study was conducted in three waves. Wave I of the NNHSF was conducted from August through December 1987. Wave II of the NNHSF was conducted from July through November 1988, approximately 12 months after the completion of Wave I. Wave III, the final wave of the study, began in January 1990 and ended in April 1990. The Office of the Assistant Secretary for Planning and Evaluation (ASPE) collaborated with NCHS and NIA in conducting Wave III of the NNHSF.

The 1985 NNHS collected a variety of information about long-term care facilities and their residents. Data were collected on a sample of patients who were current residents at the time of contact with the facility as well as a sample of discharges that occurred 12 months prior to the facility contact. There were 5,243 current residents and 6,023 discharges. For the current residents, detailed information was collected regarding dependence in activities of daily living, functional impairments, diagnoses, the receipt of services, cognitive and emotional status, charges, source of payments, history of nursing home use and a number of other topics of considerable prognostic significance. For the discharged residents, detailed information was obtained regarding diagnoses and services, source of payments, nursing home and hospital use prior to the sampled nursing home stay, hospitalization during the sample stay, and nursing home re-admissions subsequent to the sample stay. To supplement the current and discharged resident components, the 1985 NNHS included a new component—the Next-of-Kin (NOK). The NOK, using a Computer-assisted Telephone Interviewing (CATI) system, was designed to collect information about current and former nursing home residents that is not generally available from patient records or other sources in the nursing home.

The NNHSF obtains additional information on a portion of the residents for whom a Current Resident Questionnaire (CRQ) or a Discharged Resident Questionnaire (DRQ) was completed. The Wave I follow-up cohort is comprised of two types of cases. All cases with a completed NOK interview who were not known to be deceased at the time of the NOK interview were included. Cases, who were eligible for the NOK and who were not known to be deceased but for whom a completed NOK interview was not obtained, were also included in Wave I. Interviews were completed for 6,001 subjects. At the time of contact, 1,998 subjects were found to be deceased.

The NNHSF Wave II obtained additional information on those surviving subjects for whom a Wave I interview had been completed. In addition, eligible Wave I subjects were also included for whom, through interviewer error, no Wave I interview attempt was made. Wave II interviews were completed for 3,868 subjects. At the time of contact, 723 subjects were found to be deceased.

Information on the vital status of the subject (alive/deceased) at the time of the Wave II interview and the response status of the Wave II questionnaire (completed/not completed) was used to identify subjects eligible for a Wave III interview. Ultimately, a total of 3,121 subjects alive at the time of the Wave II interview and for whom some information was obtained during Wave II were deemed eligible. Wave III interviews were completed for 3,041 subjects.

In September 1994, the National Nursing Home Survey Follow-up Mortality Public-Use Data Tape was released, covering the years 1984–90. It contains the multiple cause-of-death information for 6,507 subjects from the NNHSF found to be deceased after linking and matching of files with the National Death Index. Information on the mortality tape includes the date of death, region of occurrence and residence, etc. All NNHSF tapes include a patient identification number common across files to allow linkage among them.

Data Availability: Public Use data tapes for each wave and the mortality tape are available through the National Technical Information Office (NTIS), NACDA, and the ICPSCR at the University of Michigan (see Appendix). The 1985 survey tape includes eight files: the facility questionnaire, nursing staff questionnaire, current resident questionnaire, discharged resident questionnaire, expense questionnaire, nursing staff sampling list, current resident sampling list, discharged resident sampling list. The next-of-kin questionnaire is available on a separate tape.

Linked Data: The National Nursing Home Survey Followup (NNHSF) from survey years 1985–90 is being linked to the NCHS National Death Index (1985–2002) and to SSA-RSDI (1962–2003) and SSA-SSI (1974–2003). Data are available in the NCHS Research Data Center in 2005.

Reports: A detailed description of the design and operation of the NNHSF is available in Vital and Health Statistics, Series 1, No. 30.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: [www.cdc.gov/nchs/products/elec\\_prods/subject/nnhsf.htm](http://www.cdc.gov/nchs/products/elec_prods/subject/nnhsf.htm)

## National Survey of Ambulatory Surgery (NSAS)

Sponsor: National Center for Health Statistics

Purpose: The National Survey of Ambulatory Surgery (NSAS) is the only national study of ambulatory surgical care in hospital-based and freestanding ambulatory surgery centers (ASCs). The NSAS was first conducted from 1994 to 1996, but it was discontinued due to lack of resources. The NSAS was conducted again in 2006.

Efforts are now underway to include ambulatory surgery centers in the National Hospital Ambulatory Medical Care Survey (NHAMCS). Hospital-based ASCs will be added to the scope of the NHAMCS beginning in 2009, and freestanding ASCs will be added in 2010. The [NHAMCS](#) website provides more information on the efforts.

Data for the NSAS was collected for approximately 52,000 ambulatory surgery cases from a nationally representative sample of hospital-based and freestanding ambulatory surgery centers. These data are used for a variety of planning, administrative and evaluation activities by government, professional, scientific, academic, and commercial institutions, as well as by private citizens.

After nearly ten years of being out of the field, the NSAS data collection instruments were updated to reflect the changing environment in ambulatory surgery. Outside experts from the American College of Surgeons, American Association for the Accreditation of Ambulatory Surgery Facilities (AAAASF), American Association of Ambulatory Surgery Centers (AAASC), the American Academy of Ophthalmology (AAO), the American Hospital Association (AHA), the Society for Ambulatory Anesthesiologists (SAMBA), the Federation of American Hospitals (FAH), the Joint Commission of Accreditation of Healthcare Organizations (JCAHO), the Federated Ambulatory Surgery Association (FASA), and the American College of Obstetrics and Gynecology (ACOG) provided input into updating the data collection tools used for NSAS.

Many of the new questions that were asked in the 2006 NSAS were about the facility in which the ambulatory surgery is being conducted. Each sampled facility was asked to complete a Facility Questionnaire and return it in the mail.

Data Availability: Data files for the 1994-96 and 2006 NSAS can be downloaded from the NCHS Internet site.

For more information:

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 301-458-4321

Website: <http://www.cdc.gov/nchs/nsas.htm>

## **National Vital Statistics System (Vital Statistics)**

Sponsor: National Center for Health Statistics

Data Collection Agency/Organization: National Center for Health Statistics

Purpose: Through the National Vital Statistics System, the National Center for Health Statistics collects and publishes data on births, deaths, and prior to 1996, marriages and divorces occurring in the United States based on U.S. Standard Certificates. The Division of Vital Statistics obtains information on births and deaths from the registration offices of each of the 50 States, New York City, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, Guam, America Samoa, and Northern Mariana Islands. Geographic coverage for births and deaths has been complete since 1933. Demographic information on the death certificate 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.

For more information on the mortality data files, Kung HC, Hoyert DL, Xu JQ, Murphy SL. Deaths: final data for 2005. National vital statistics reports; vol 56, no 10. Hyattsville, MD: National Center for Health Statistics. 2008.

For more information:

Contact: Mortality Statistics Branch

Email: [nchsquery@cdc.gov](mailto:nchsquery@cdc.gov)

Phone: 1-866-441-NCHS (6247)

Website: <http://www.cdc.gov/nchs/deaths.htm>

## **Health Data Interactive (formerly Trends in Health and Aging)**

Sponsor: National Institute on Aging, National Center for Health Statistics

Data Collection Agency/Organization: National Center for Health Statistics

The National Center for Health Statistics has combined two interactive, web-based applications, Health Data for All Ages and Trends in Health and Aging, to create a new site. Most of the table topics and many of the previous tables will continue to be featured on HDI. Where possible, age groups have been expanded and the depth of demographic detail increased. In some cases, we offer two tables on the same topic, with one table featuring detailed data on subpopulations and another table providing longer trend data.

Recently Updated Tables: The following tables have recently been updated or added to HDI:

- Amount of assistance with activities of daily living, among nursing home residents: US, 2004 (Source: NNHS)
- Difficulty in physical functioning among Medicare beneficiaries, ages 65+: US, 1992-2006 (Source: MCBS)
- Difficulty performing activities of daily living among Medicare beneficiaries, ages 65+: US, 1992-2006 (Source: MCBS)
- Difficulty performing instrumental activities of daily living among Medicare beneficiaries, ages 65+: US, 1992-2006 (Source: MCBS)
- Functional limitations among Medicare beneficiaries, ages 65+: US, 1992-2006 (Source: MCBS)
- Hospital discharges by first-listed diagnosis: US, 1990-2006 (Source: NHDS)
- Nursing home residents: US, 1977-2004 (Source: NNHS)

Future Plans: As we continue our transition from Trends in Health and Aging and Health Data for All Ages, we will add the following tables to HDI:

- Difficulty in physical functioning among Medicare beneficiaries, ages 65+: US, 1992-2005 (Source: MCBS)
- Hospital discharges by first-listed diagnosis among children: US, 1989-2006 (Source: NHDS)
- Hospital procedures, all-listed: US, 1990-2006 (Source: NHDS)
- Mortality by underlying cause, ages 18+: US/State, 2000-2005 (Source: NVSS)
- Mortality by underlying cause, ages 45+: US, 1968-1980 (Source: NVSS)
- Smoking, ages 25+, trend: US, selected years 1965-2006 (Source: NHIS)
- Vaccinations for influenza and pneumonia, ages 18+: US, selected years 1993-2006 (Source: NHIS)

## **Centers for Medicare & Medicaid Services**

### **Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surveys**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Westat

Purpose: The purpose of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Surveys are to measure experiences of enrollees in their health and prescription drug plans and with fee-for-service Medicare and to measure the experiences of patients in hospitals and other provider settings such as nursing homes, End-Stage Renal disease facilities, and home health agencies. The health plan surveys have been conducted annually by the Centers for Medicare & Medicaid Services (CMS). Hospital CAHPS® was implemented nationally in 2006. The other CAHPS® surveys are in different phases of development.

The primary goals of these surveys are to:

- Provide people with Medicare and the general public with information to help them make more informed choices among health plans and providers
- Help CMS identify problems and improve the quality of services
- Enhance CMS' ability to monitor the quality of care and performance of its Medicare plans and providers

For more information:

Website: <http://www.hcahponline.org/home.aspx>

## **Medicaid (Claims and Enrollment) Data**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: States

Purpose: The Centers for Medicare & Medicaid Services (CMS) works with its State partners to collect data on persons served by the Medicaid program to produce Medicaid program statistics, monitor and evaluate access and quality of care, trends in program eligibility, characteristics of enrollees, changes in payment policy, and other program-related issues.

Selected Content: Medicaid enrollment data collected include demographic and eligibility characteristics of enrolled persons such as race/ethnicity, age, and basis of eligibility. Enrollee eligibility variables include state, county and zip code of residence, although reported zip code is of varying quality. Service data include detail on utilization and Medicaid payments for each covered service.

Data Years: Selected State data are available from 1992 forward. Data for the 50 States and the District of Columbia are available from 1999 forward.

Coverage: The data include individuals enrolled in the Medicaid program and the Medicaid-covered services they receive.

Methodology: The primary data source for Medicaid statistical data is the Medicaid Statistical Information System (MSIS). The Medicaid Analytic eXtract (MAX) data, derived from MSIS, are the primary data used for research and policy analysis. Prior to 1999 the predecessor to MAX was the State Medicaid Research Files (SMRF). Detailed information on MAX and SMRF are available on the CMS website at:

[www.cms.hhs.gov/MedicaidDataSourcesGenInfo/07\\_MAXGeneralInformation.asp](http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/07_MAXGeneralInformation.asp).

MAX and SMRF files contain person-level enrollment, utilization, and expenditure data on a calendar year basis. For years prior to 1999, they are available for 25–31 States, depending on the year. These are States that chose to participate voluntarily in electronic data submission or the Medicaid Statistical Information System (MSIS). MSIS participation was mandated for all States beginning with 1999. Calendar year MAX and SMRF files are created from the quarterly MSIS files that State Medicaid agencies submit to CMS. MAX and SMRF include one file with enrollment information (Personal Summary File) and four service files (Inpatient, Long Term Care, Prescription Drug, and Other Services) for each year of data. In MAX and SMRF service files, interim claims (originals, voids, and positive or negative adjustments) have been combined so that the records represent final action “events” to the extent possible. While MAX and SMRF data have undergone extensive edit checks, Medicaid programs and data quality vary across States.

MSIS is the basic source of State-reported eligibility and claims data on the Medicaid population, and their characteristics, utilization, and payments. Beginning in FY 1999, 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 MSIS. Prior to FY 1999, 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 HCFA-2082 report was produced by the Health Care Financing Administration (the predecessor to CMS).

Research Identifiable Files (RIFs): MAX and SMRF are research identifiable files. Person-level identifiable information is included in these Medicaid enrollment and utilization files. These data are protected under the Privacy Act of 1974 and other applicable Federal government rules and regulations. A Privacy Act Systems of Records notice (No. 09-70-6001), published in the Federal Register, identifies allowable uses of the data if potential benefits outweigh the risk of disclosure. Research is identified as an allowable use. CMS employs strict security measures to safeguard individual privacy. CMS data release policies seek to ensure that files containing physician and/or beneficiary identifiers are used only when necessary and in accordance with disclosure provisions of the Privacy Act. Researchers need to submit to CMS a data request package containing a written request, study protocol, evidence of funding, and a Data Use Agreement (DUA). The package is reviewed by the CMS Privacy Board. If CMS approves the data file request, researchers need to pay the cost incurred in the processing of data. Only the minimum data necessary for the research study are released to an approved user. Further information is available at the following CMS web addresses:

<http://www.cms.hhs.gov/PrivProtectedData/> and  
<http://www.cms.hhs.gov/cmsforms/downloads/cms-r-0235.pdf>.

The CMS Medicaid enrollment and claims data contains information for Medicaid eligibles who actually do enroll in their State's Medicaid program. CMS Medicaid files cannot, therefore, be used to study individuals who are eligible but not enrolled.

For more information:

Contact: CMS

Website: <http://www.cms.hhs.gov/home/medicaid.asp>

Contact: Research Data Assistance Center

Phone: 1-888-973-7322

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Internet: [www.resdac.umn.edu](http://www.resdac.umn.edu)

## **Medicare (Fee-for-Service (FFS) Claims, Part D Events, and Parts A/B/C/D Enrollment) Data**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Centers for Medicare & Medicaid Services

Purpose: The Centers for Medicare & Medicaid Services (CMS) collects and synthesizes Medicare enrollment, FFS spending, and FFS claims data to monitor and evaluate access to and quality of care, trends in utilization, changes in payment policy, and other program-related issues.

Selected Content: Data include claims information for FFS services furnished to Medicare beneficiaries and Medicare enrollment data. Claims data include type of service, procedures, diagnoses, dates of service, and claim-reported amounts for Medicare spending, for other primary payers, and potential beneficiary liability for deductible and coinsurance. Enrollment data include date of birth, sex, race/ethnicity, state and county of residence, and reason for entitlement.

Data Years: Some data files are available as far back as 1987, but CMS no longer provides technical support for files with data prior to 1996.

Coverage: Enrollment data are for all persons enrolled in the Medicare program (original or FFS Medicare, managed care (Part C), and Part D). FFS claims data include data for Medicare beneficiaries who filed claims or had claims filed on their behalf. Part D event data include selected information extracted from claims by Part D plans and submitted to CMS.

Methodology: Analysis of the Medicare program and its impact on beneficiaries should start by first focusing on the targeted benefit and the population at risk.

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. A modification to the Denominator File to include information on enrollees’ Part D status, including deemed/low income cost sharing, type of Part D plan, state-reported dual eligibility, and retiree drug subsidy, is now (August 2008) being prepared.

On occasion additional information on enrollees in plans (either managed care organization or Part D plans) is needed. The Management Integrated Information Repository (MIIR) contains data on beneficiaries who are currently enrolled or have ever been enrolled in a Managed Care

Organization (MCO) under contract with CMS to provide A/B services and in a Prescription Drug Plan (PDP) or Medicare Advantage Part D (MA PD) Plan to provide prescription drug coverage. The data are individually beneficiary based. 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/Part D enrollment information such as dates of membership and MCO/PBP contract number. The MIIR File is used to identify the exact MCO/PDP/MA PD in which beneficiaries were enrolled.

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. This file, like the Denominator File is derived from the Enrollment Data Base (EDB).

Most analyses of the FFS sector involve claims for A/B services. 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 the National Claims History (NCH) files, the Standard Analytic Files (SAF), Medicare Provider and Analysis Review (MedPAR) files, Medicare enrollment files, and various other files.

The National Claims History (NCH) 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, and the Durable Medical Equipment SAF, all of which are potentially available as 100 % samples. In addition, representative 5% Sample Beneficiary SAF's are created for inpatient hospital, SNF, home health, hospice, outpatient, physician/supplier, and DME.

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.

Issues Affecting Interpretation: Because Medicare managed care organization do not file claims, files based only on FFS claims data will exclude care for persons enrolled in Medicare managed care programs. In addition, to maintain a manageable file size, some files are based on a sample of enrollees, rather than on all Medicare enrollees. Coding changes and interpretation

of Medicare coverage rules have also changed over the life of the Medicare program. In addition, the availability of Part D utilization in the form of abstracts of the Part D claims called prescription drug events has been limited to payment purposes only, but efforts are being made currently (July 2008) to develop plans for disseminating the data for research purposes within the confines of the May 28, 2008 Final Rule regarding Part D Claims Data.

For more information:

Contact: Research Data Assistance Center

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Phone: 1-888-973-7322

Website: [www.resdac.umn.edu](http://www.resdac.umn.edu)

## **Medicare Current Beneficiary Survey (MCBS)**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Westat

**Purpose:** The Medicare Current Beneficiary Survey (MCBS) produces nationally representative estimates of health status, health care use and expenditures, health insurance coverage, and socioeconomic and demographic characteristics of Medicare beneficiaries. It is used to estimate expenditures and sources of payment for all services used by Medicare beneficiaries, including co-payments, deductibles, and noncovered services; to ascertain all types of health insurance coverage and relate coverage to sources of payment; and to trace processes over time, such as changes in health status, spending down to Medicaid eligibility, and the effects of program changes.

**Selected Content:** The survey collects 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).

**Data Years:** The first round of interviewing was conducted from September through December 1991, and the survey has been continuously in the field since then. The data are designed to support both cross-sectional and longitudinal analyses.

**Coverage:** The MCBS is a continuous survey of a nationally representative sample of aged, institutionalized, and disabled Medicare beneficiaries.

**Methodology:** 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 caregivers 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 those who enter Medicare during the year and those who died.

Sample Size and Response Rate: Each fall, about one-third of the sample is retired and roughly 6,000 new sample persons are included in the survey—the exact number chosen is based on projections of target samples of 12,000 persons with 3 years of cost and use information distributed appropriately across the sample cells. In the community, response rates for initial interviews range in the mid- to high 80s; once respondents have completed the first interview, their participation in subsequent rounds is 95 percent or more. In recent rounds, data have been collected from approximately 15,000 to 19,000 beneficiaries, with the peaks occurring in fall rounds. Roughly 90 percent of the sample is made up of persons who live in the community, with the remaining persons living in long-term care facilities. Response rates for facility interviews approach 100 percent.

Issues Affecting Interpretation: Because only Medicare enrollees are included in the survey, the survey excludes a small proportion of persons age 65 and over who are not enrolled in Medicare, which should be noted when using the MCBS to make estimates of the entire population age 65 and over in the United States.

References:

Adler GS. A profile of the Medicare Current Beneficiary Survey. *Health Care Finance Rev*;15(4):153-163. 1994.

Lo A, Chu A, Apodaca R. Redesign of the Medicare Current Beneficiary Survey Sample, Rockville, MD: Westat, Inc., 2003, found at <http://www.amstat.org/sections/SRMS/Proceedings/y2002/Files/JSM2002-000662.pdf>.

For more information:

Contact: MCBS

Website: <http://www.cms.hhs.gov/mcbs/>

Research Data Assistance Center

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Phone: 1-888-973-7322

Website: [www.resdac.umn.edu](http://www.resdac.umn.edu)

## **Medicare Health Outcomes Survey (HOS)**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: National Committee for Quality Assurance

**Purpose:** The Medicare Health Outcomes Survey (HOS) is the first outcomes measure used in Medicare managed care. The goal of the Medicare HOS program is to gather valid, reliable, and clinically meaningful health status data in Medicare managed care for use in quality improvement activities, plan accountability, public reporting, and improving health. All managed care plans with Medicare Advantage (MA) (formerly Medicare +Choice) contracts must participate. The HOS assesses an MA Plan's ability to maintain or improve the physical and mental health of its Medicare member over time.

The Centers for Medicare & Medicaid Services (CMS), in collaboration with the National Committee for Quality Assurance (NCQA), launched the Medicare HOS in 1998. The Medicare HOS is being used as part of the effectiveness of care component of the Healthcare Effectiveness Data and Information Set (HEDIS®). This measure was initially titled Health of Seniors, and was renamed the Medicare Health Outcomes Survey during the first year of implementation. This name change was intended to reflect the inclusion of people with Medicare who are disabled and under age 65 in the sampling methodology.

The HOS measure was developed under the guidance of a Technical Expert Panel comprised of individuals with specific expertise in the health care industry and outcomes measurement. The measure includes the most recent advances in summarizing physical and mental health outcomes results and appropriate risk adjustment techniques. In addition to health outcomes measures, the HOS is used to collect height and weight to calculate body mass index, as well as the Urinary Incontinence in Older Adults, Physical Activity in Older Adults, Fall Risk Management, and Osteoporosis Testing in Older Women HEDIS® measures.

**Survey Universe:** Aged and disabled Medicare beneficiaries continuously enrolled in the same MA Plan for 6 months. Institutionalized members are eligible. Members with end stage renal disease (ESRD) are excluded.

**Research Design:** The HOS is an ongoing longitudinal, nationally representative Medicare managed care survey.

**Sample:** A random sample of Medicare beneficiaries is drawn from each participating plan and surveyed every spring (i.e., a survey is administered to a new baseline cohort, or group, each year). Two years later, the baseline respondents are surveyed again (i.e., follow-up measurement). Participating MA plans must have a minimum enrollment of 500 at the time of baseline sampling. For data collection years 1998-2006, the plan sample size was one thousand. Effective 2007, the plan sample size increased to twelve hundred. For MA Plans with 1,200 members or less, all eligible members are included in the sample for the baseline survey. Cohort One was surveyed in 1998 and was resurveyed in 2000. During the 2008 HOS administration,

Cohort Eleven was surveyed and Cohort Nine was resurveyed. Nearly two million beneficiaries have been surveyed to date.

Survey Mode: The HOS is a mail survey with telephone (CATI) follow up for nonrespondents.

Unit of Analysis: Medicare managed care (i.e., Medicare Advantage) plan.

Topics/Key Words: Medicare Advantage; Medicare Managed Care; Health Outcomes; Functional Status; Quality Improvement; Performance Measurement; Health Status.

Data Availability: Baseline Data, Cohorts 1–9 (Baseline Measurement Years 1998-2006) and 2-Year Analytic Cohort Data, Cohorts 1–7 (Follow Up Measurement Years 2000–06) are currently available. Follow up HOS data sets are available only for Cohorts 1-5 (2000-2004). Data for 2007 Cohort 10 Baseline and 2005-2007 Merged Analytic Cohort 8 will be available later in 2008.

Linked Data: Since 2003, the HOS instrument includes the Healthy Days Measures from the Centers for Disease Control and Prevention's (CDC's) Behavioral Risk Factor Surveillance System (BRFSS). The inclusion of these questions allows a link between HOS and BRFSS results.

In addition, data from HOS Cohorts 1-4 (measurement years 1998-2004) are currently linked with data from NCI's Surveillance, Epidemiology, and End Results (SEER) cancer registry.

Data Dissemination: After the administration of each baseline cohort, Quality Improvement Organizations (QIOs) receive beneficiary level data files for each MA plan in their respective States. Beneficiary level performance measurement data are disseminated to all participating QIOs and made available to participating MA Plans after the follow up measurement for each cohort.

Baseline and Merged 2-Year Analytic HOS public use files (PUFs) are available and can be found at <http://www.hosonline.org/surveys/hos/hosdata.aspx> to facilitate additional research. Medicare HOS limited data sets (LDSs) and research identifiable files (RIFs) are also available to researchers with a signed data use agreement with CMS.

Reports: Each year a Baseline Report and a Performance Measurement Report are disseminated to each MA Plan participating in the Medicare HOS. Each participating MA Plan receives plan specific baseline and/or performance measurement reports, which present results for their plan, the State total, and the HOS national total. Additionally, each State's QIO receives State specific baseline and/or performance measurement reports, which present results for all plans in their State, the State total, and the HOS national total.

Research utilizing Medicare HOS data has also resulted in a number of technical reports, manuals, and peer-reviewed articles. More information on available HOS publications can be found at <http://www.hosonline.org/surveys/hos/hospublications.aspx>.



For more information:

Chris Haffer, Ph.D.  
Director, Medicare HOS Program  
Phone: 410-786-8764  
Email: [chris.haffer@cms.hhs.gov](mailto:chris.haffer@cms.hhs.gov)

Sonya Bowen, MSW  
Program Administrator  
Phone: 410-786-5609  
Email: [sonya.bowen@cms.hhs.gov](mailto:sonya.bowen@cms.hhs.gov)

A general overview of the Medicare HOS program is available at the CMS HOS web address (<http://www.cms.hhs.gov/hos/>). For comprehensive information on the history and progress of the Medicare HOS program, including the full spectrum of available data and reports, see <http://www.hosonline.org/>.

## **Minimum Data Set (MDS)**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Long term care facilities

Purpose: The Minimum Data Set (MDS) is a uniform set of elements extracted from the Resident Assessment Instrument (RAI), which is a standardized tool for assessing the functional capacity of residents of long term care facilities. On December 23, 1997, the Centers for Medicare & Medicaid Services (CMS) published the final rule, which established the guidelines for the use of the dataset and specified the data elements included in the assessment instrument. Long term care facilities are required to complete and transmit MDS data to the designated State agency for all residents as a condition of participation in the Medicare and Medicaid programs. Automated transmission of all MDS data was required beginning in 1998. CMS began collecting MDS data in a national database in July 1998.

The Long Term Care Minimum Data Set (MDS) is a standardized, primary screening and assessment tool of health status, which forms the foundation of the comprehensive assessment for all residents of long-term care facilities certified to participate in Medicare or Medicaid.

The MDS contains items that measure physical, psychological, and psycho-social functioning. The items in the MDS give a multidimensional view of the patient's functional capacities, and can be used to present a nursing home's profile. The MDS can be linked to Medicare claims data and to survey and certification facility data to build a number of resident-level and facility-level analytical datasets. The MDS also now plays a key role in the Medicare and Medicaid reimbursement system and in monitoring the quality of care provided to nursing facility residents.

For more information:

Contact: Research Data Assistance Center

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Phone: 888-973-7322

Website: [www.resdac.umn.edu](http://www.resdac.umn.edu)

## **Outcome and Assessment Information Set (OASIS)**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Home health agencies

Purpose: The Home Health Outcome and Assessment Information Set (OASIS) contains data items that were developed for measuring patient outcomes for the purpose of performance improvement in home health care. Medicare certified home care agencies are required to conduct patient-specific comprehensive assessments at specified time points. The assessments for Medicare and Medicaid patients must contain the OASIS data elements. Some data elements are also used to determine Home Health Prospective Payment System reimbursement. This data is used to prepare OASIS Outcome reports for agency performance improvement as well as the reports posted on Home Care Compare. The data are collected at start of care, 60-day follow-ups, and discharge (and transfer to and from an inpatient stay).

OASIS data items address sociodemographic, environmental, support system, health status, functional status, and health service utilization characteristics of the patient.

Data Availability: OASIS data from CMS available since 1999.

For more information:

Contact: Research Data Assistance Center

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Phone: 888-973-7322

Website: <http://www.resdac.umn.edu/>

## **Section 723 Chronically Ill Medicare Beneficiary Research and Data Initiative (Section 723)**

Sponsor: Centers for Medicare & Medicaid Services

Data Collection Agency/Organization: Centers for Medicare & Medicaid Services

Purpose: Section 723 of the Medicare Modernization Act of 2003 called for the establishment of a chronic condition data warehouse (CCW) to support the study of chronically ill Medicare beneficiaries and of cost effective and quality improvement options to treat these beneficiaries. The first phase of this work was done at the Iowa Foundation for Medical Care. The next phase will be performed by Buccaneer.

The data warehouse supports:

- Studies for improving the quality of care for chronically ill Medicare beneficiaries
- Studies for reducing the cost of care
- Integration of existing datasets
- Identification of new data needs for research
- Consultations with experts in the fields of care for the chronically ill

Data stored in the warehouse have the following features:

- Unique patient ID linked to all CMS program data creating person-level view of data
- Data extraction tools that support
  - Accessing data by chronic conditions
  - Complex customized research data requests related to chronic illness

Data included in the warehouse will be a 5 percent sample of all Medicare beneficiaries from CY 1999 through 2004 (about 2.5 million people), and a 100 percent sample from starting in CY 2005 linking:

- Medicare enrollment data
- Medicare claims
- Medicaid claims
- MDS (nursing facility), OASIS (home health, IRF PAI (rehabilitation facility), and SB (swing bed) assessments
- Beneficiary data and data collected from relevant surveys (e.g., MCBS)

Future Activities:

- Expand data sources to include Part D enrollment and drug events
- Enhance data access tools
- Enhance technical support services

For more information:

Contact: Research Data Assistance Center

Email: [resdac@umn.edu](mailto:resdac@umn.edu)

Phone: 1-888-973-7322

Website: <http://www.resdac.umn.edu/>

## **National Institutes of Health**

### **National Cancer Institute**

#### **Surveillance, Epidemiology, and End Results (SEER)**

Sponsors: National Cancer Institute, Centers for Disease Control and Prevention

Data Collection Agency/Organization: National Cancer Institute

Purpose: The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) is an authoritative source of information on cancer incidence and survival in the United States. SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 26 percent of the U.S. population. SEER coverage includes 23 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Hawaiian/Pacific Islanders. (Details are provided in the table: Number of Persons by Race and Hispanic Ethnicity for SEER Participants.) The SEER Program registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status. The SEER Program is the only comprehensive source of population-based information in the United States that includes stage of cancer at the time of diagnosis and patient survival data.

SEER began collecting data on cancer cases on January 1, 1973, in the States of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit and San Francisco-Oakland. In 1974–75, the metropolitan area of Atlanta and the 13-county Seattle-Puget Sound area were added. In 1978, 10 predominantly black rural counties in Georgia were added, followed in 1980 by the addition of American Indians residing in Arizona. Three additional geographic areas participated in the SEER program prior to 1990: New Orleans, Louisiana (1974–77, rejoined 2001); New Jersey (1979–89, rejoined 2001); and Puerto Rico (1973–89). The National Cancer Institute also funds a cancer registry that, with technical assistance from SEER, collects information on cancer cases among Alaska Native populations residing in Alaska. In 1992, the SEER Program was expanded to increase coverage of minority populations, especially Hispanics, by adding Los Angeles County and four counties in the San Jose-Monterey area south of San Francisco. In 2001, the SEER Program expanded coverage to include Kentucky and the remaining counties in California (Greater California); in addition, New Jersey and Louisiana once again became participants. For the expansion registries (Kentucky, Greater California, New Jersey, and Louisiana), NCI funds are combined with funding from the Centers for Disease Control and Prevention (CDC) through the National Program of Cancer Registries and with funding from the States.

NCI staff work with the North American Association of Central Cancer Registries (NAACCR) to guide all State registries to achieve data content and compatibility acceptable for pooling data and improving national estimates. The SEER team is developing computer applications to unify cancer registration systems and to analyze and disseminate population-based data. Use of surveillance data for research is being improved through Web-based access to the data and

analytic tools, and linking with other national data sources. For example, a new Web-based tool for public health officials and policymakers, State Cancer Profiles, provides a user-friendly interface for finding cancer statistics for specific States and counties. This website is a joint project between NCI and CDC and is part of the Cancer Control PLANET website, which provides links to comprehensive cancer control resources for public health professionals.

The SEER Program is considered the standard for quality among cancer registries around the world. Quality control has been an integral part of SEER since its inception. Every year, studies are conducted in SEER areas to evaluate the quality and completeness of the data being reported.

For more information:

Contact: Cancer Statistics Branch, National Cancer Institute

Email: [seerweb@imsweb.com](mailto:seerweb@imsweb.com)

Phone: 301-496-8510

Website: <http://seer.cancer.gov/>

## **National Institute on Aging**

### **Health and Retirement Study (HRS)**

Sponsor: National Institute on Aging

Data Collection Agency/Organization: Institute for Social Research, University of Michigan

Purpose: The HRS is designed to provide a uniquely rich, nationally representative longitudinal dataset for the community of scientific and policy researchers who study the health, economics, demography, sociology, and psychology of aging. The fact that data from all these domains are collected in the same survey creates opportunities for new types of interdisciplinary research. In addition, the HRS provides researchers with the capacity to study the effects of expected and unexpected changes in policy, the economy or the society in a timely way, taking advantage of the many “natural experiments” that occur over time. For example, what is the effect of the stock market decline after 2001 on savings and retirement decisions?

The objectives of the study are to:

- Explain the antecedents and consequences of retirement
- Examine the relationship between health, income, and wealth over time
- Examine life cycle patterns of wealth accumulation and consumption
- Monitor work disability
- Provide a rich source of interdisciplinary data, including linkages with administrative data
- Examine how the mix and distribution of economic, family and program resources affect key outcomes, including retirement, dissaving, health declines, and institutionalization

Study Design:

The design has these characteristics:

- National panel study
- Initial sample of over 12,600 persons in 7,600 households
- Current total sample of over 22,000 persons in 13,100 households
- Oversamples of Hispanics, Blacks, and Florida residents
- Baseline: in-home, face-to-face in 1992 for the 1931–41 birth cohort (and their spouses, if married, regardless of age); in 1998 for 1924–30 (CODA: Children of the Depression Age) and 1942–47 (War Babies) birth cohorts; in 2004 for 1948–53 (Early Boomers) cohort
- Follow-ups by telephone every second year, with proxy interviews after death. Beginning in 2006, half the sample will have enhanced face-to-face follow-ups that will include the collection of physical measures and biomarkers.

HRS provides a research data base that can simultaneously support continuous cross-sectional descriptions of the U.S. population over the age of 55, longitudinal studies of a given cohort over a substantial period of time (up to 18 years by 2010 for the original HRS cohort, following them from age 51–61 to age 69–79) and research on cross-cohort trends. By 2010 the HRS will be able to support cross-cohort comparisons of trajectories of health, labor supply, or wealth accumulation for persons who entered their 50s in 1992, 1998, and 2004.



#### Questionnaire Topics:

Health and cognitive conditions and status, including physical measures  
Retirement plans and perspectives  
Psycho-social well-being  
Attitudes, preferences, expectations, and subjective probabilities  
Family structure and transfers  
Employment status and job history  
Job demands and requirements  
Disability  
Demographic background  
Housing  
Income and net worth  
Health insurance and pension plans  
Experimental modules

#### Links with Administrative Data:

Employer Pension Plans  
National Death Index  
Social Security Administration earnings and (projected) benefits data; W-2 self-employment data  
Medicare and Medicaid files

#### Study of Assets and Health Dynamics among the Oldest Old (AHEAD):

AHEAD is a national panel study with an initial sample of 7,447 respondents aged 70+ (and their spouses, if married, regardless of age). There is a supplemental sample of respondents aged 80+ from the Medicare Master Enrollment File. Wave 1 data collection was completed in 1994 and Wave 2 in 1996. Since 1998, AHEAD respondents have been contacted as part of a joint data collection effort with the HRS.

AHEAD provides detailed coverage of the joint dynamics among health (physical, cognitive, and functional), dementia, economic and family resources, and care arrangements. The AHEAD study provides data on the interplay of resources and late life health transitions, including: the cost of illness borne by the family; differences in how resources are used to offset cognitive, physical, and functional losses; the effectiveness of various care arrangements in preserving function and delaying institutionalization; the extent to which transfers from kin buffer the assets of older persons and slow transitions to late life impoverishment; and the extent and mechanisms for dissaving and Medicaid spend down.

#### Aging, Demographics, and Memory Study (ADAMS):

One component of the data system is the use of the HRS survey as a sampling frame to support targeted sub-studies. HRS has been used as a sampling frame for the Aging, Demographics, and Memory Study (ADAMS) supplement on dementia. This supplement to the HRS conducts a field assessment of a sample of about 930 HRS panel members aged 75+ to clinically assess their dementia status and dementia severity. In combination with multiple waves of HRS data on cognitive performance, functional status, medical history, and socioeconomic measures, a

probability function of dementia will be estimated for HRS participants. The addition of this imputed dementia status to the ongoing panel will provide a unique resource for the study of the costs of dementia to the individual and his family. In years 17–22, there is a plan to convert the ADAMS sample into a dementia incidence study, exploiting the highly detailed baseline information collected previously.

The HRS is also used as a sampling frame for mail surveys on other special topics including consumption and time use, prescription drug use and the impact of Medicare Part D, parents' human capital investments in children, and diabetes management by self-reported diabetics; and for a number of experimental topics using Internet interviewing. Because a lot is known about HRS respondents, the HRS provides an efficient frame from which to draw probability samples of relatively rare groups such as the cognitively impaired or diabetics. In addition, research using results from a sub-study can make use of all the information collected in the HRS from waves prior to the sub-study to deal with issues involving initial conditions and selectivity and also automatically provides prospective follow-up information in the core longitudinal survey.

#### Comparable International Data:

One aspect of the HRS data system, which is not formally part of the HRS project, is the emergence of a set of longitudinal studies of aging in other countries that have been consciously designed to produce data as comparable as possible to that collected by the HRS. These studies include English Longitudinal Study of Ageing (ELSA), Survey of Health, Ageing and Retirement in Europe (SHARE), in 11 countries in Europe, and Mexican Health and Aging Study (MHAS). Israel is planning to enter SHARE, and South Korea is planning to launch a longitudinal study of aging as well. The HRS has actively collaborated with all of these studies, providing both scientific and technical help.

#### Future Plans:

During years 17–22 of the HRS, the plan is to continue with core data collection on the steady-State design with biennial longitudinal data collection, including the addition of a new cohort in 2010; to continue the aims of the ADAMS supplement on dementia; to add the collection of biological samples, including DNA, physical performance measures and psychosocial content; and to continue the use of mail surveys on special topics, including continued longitudinal measurement of consumption and time use for about half of the sample, as a low-cost complement to the core data collection.

#### Data Availability:

All publicly available data may be downloaded after registration from <http://hrsonline.isr.umich.edu>. Early Release data files are typically available within 3 months of the end of each data collection, with the Final Release following at 24 months after the close of data collection activities. Files linked with administrative data are released only as restricted data through an application process, as outlined on the HRS website.

#### Bibliography:

The HRS bibliography of nearly a thousand publications is on-line at [http://hrsonline.isr.umich.edu/papers/sho\\_papers.php?hfyle=bib\\_all](http://hrsonline.isr.umich.edu/papers/sho_papers.php?hfyle=bib_all). To search for a specific

publication or topic in the bibliography, click on the link for the Dynamic Bibliography or go to <http://hrsonline.isr.umich.edu/biblio/index>.

For more information:

Contact: Robert J. Willis or David R. Weir, Co-Directors

Email: [hrsquest@isr.umich.edu](mailto:hrsquest@isr.umich.edu)

Phone: 734-936-7261

Website: <http://hrsonline.isr.umich.edu/>

## **National Long Term Care Survey (NLTC)**

Sponsors: National Institute on Aging & Office of the Assistant Secretary for Planning and Evaluation

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: The 1982, 1984, 1989, 1994, 1999, and 2004 National Long Term Care Surveys (NLTC) are surveys of the entire Medicare enrolled aged population with a particular emphasis on the aged who are functionally impaired. The samples drawn from aged Medicare beneficiary enrollment files are nationally representative of both community and institutional residents. As sample persons are followed through the Medicare record system, virtually 100 percent of cases can be longitudinally tracked so that declines, as well as increases, in disability may be identified as well as exact dates of death. NLTC sample persons are followed until death and are permanently and continuously linked to the Medicare record system from which they are drawn. Linkage to the Medicare Part A and B service use records extends from 1982 to 2004, so that detailed Medicare expenditures and types of service use may be studied. Medicare file updates are received on an annual basis.

Description and Study Design: The NLTC was designed to measure the point prevalence of chronic (90 days or more) disability in the U.S. elderly Medicare enrolled population and changes (both improvement and incidence) in chronic disability (and institutionalization) over time. In addition, it measures the relation of such changes to health status (self report and Medicare file diagnosis) vital States and to detailed records of Medicare service use and costs.

The 1982 NLTC sample was randomly drawn from Medicare administrative files. Over 35,000 persons over age 65 who were Medicare eligible and alive on April 1, 1982, were drawn. A two-stage procedure identified chronically disabled persons for household interviews. This identified 6,393 persons with at least one chronic (duration of 90 days or more) impairment in seven Instrumental Activities of Daily Living (IADL) or nine Activities of Daily Living (ADL). Since residence was confirmed in the screening, institutional residents can be separated from the nondisabled community residents who reported no ADL or IADL impairments. Chronically disabled community residents were interviewed at home to assess health, functioning, social, economic, and other factors. No institutional interviews were conducted in 1982.

The 1984 NLTC had a complete longitudinal design with an institutional component. All persons surviving to 1984 who, in 1982 either reported chronic disability or were in institutions, were re-interviewed in 1984 with either a detailed community or institutional questionnaire. Consequently, not only disabled survivors were tracked, but also previously disabled survivors who had become nondisabled. In addition, community residents screened in 1982, and who were not chronically disabled, were rescreened with the same instrument in 1984 to determine disability incidence. A new sample who became age 65 between the 1982 and 1984 surveys and who survived to 1984 was drawn from Medicare files and screened. The 1984 sample provides estimates of 2-year disability and mortality rates for a longitudinally followed population and representative cross-sectional prevalence estimates of disability and institutional residence of the

total 1984 U.S. elderly Medicare enrolled population 65 and over. The Next-of-Kin survey was conducted along with this wave.

The 1989 NLTCS was designed to estimate disability and institutionalization rates for the 1982–84 and 1984–89 periods. Persons in either 1982 (or 1984) who were institutionalized, or who received a detailed community survey and survived to 1989, were re-interviewed—either in the community or in an institution. An “aged-in” sample who became 65 between 1984 and 1989 and survived to 1989 were selected and screened for disability. All nondisabled persons over age 75 (in 1989) from the 1984 sample were automatically screened in 1989. The Informal Caregiver survey was conducted with this wave.

The 1994 NLTCS was designed to precisely estimate disability and institutionalization rates for the 1982–94 period. Persons in either the 1982, 1984, or 1989 surveys who received a detailed interview in the community or an institution and survived to 1994 were re-interviewed, wherever they might reside. An aged-in sample who became 65 between 1989 and 1994 and survived to 1994 were selected and screened for disability. Furthermore, in 1994, a supplementary group of community interviews were conducted with nondisabled persons (the healthy supplement). This group will be followed like all other detailed interview persons in subsequent years. Further, a group age 95+ (N = 540) were also drawn in 1994 and screened for disability. This supplemental sample greatly increases the precision associated with estimates of the extremely aged.

The 1999 NLTCS was designed to precisely estimate disability and institutionalization rates for the 1982–99 period. In 1999, NLTCS employed, for the first time, Computer-assisted Personal Interviewing (CAPI). Persons in either the 1982, 1984, 1989, or 1994 surveys who received a detailed interview (in the community or in an institution) and survived to 1999 were reinterviewed if they resided in the United States. An “aged-in” sample who became 65 between 1994 and 1999 and survived to 1999 were selected and screened for disability. A supplementary group was selected for detailed interviews as part of the “healthy subsample.” This group, some of whom were interviewed in 1994, will be followed like other detailed interview groups in future survey years. Also, a group of persons age 95+ (N = 600) were selected in 1999, enabling precise estimates of the extreme aged population. The Informal Caregiver and the Next-of-Kin surveys were conducted with this wave.

The 2004 NLTCS was designed to precisely estimate disability and institutionalization rates for the 1982–2004 period. In 2004, NLTCS again employed CAPI. Persons in either the 1982, 1984, 1989, 1994, or 1999 surveys who received a detailed interview (in the community or in an institution) and survived to 2004 were reinterviewed if they resided in the United States. An “aged-in” sample that became 65 between 1999 and 2004 and survived to 2004 were selected and screened for disability. A supplementary group was selected for detailed interviews as part of the “healthy subsample.” This healthy subsample, some of whom were interviewed in 1994 and 1999, will be followed like other detailed interview groups in future survey years. Also, a group of persons aged 95+ (N = 1,584) were selected in 2004, enabling precise estimates of the extreme aged population. A subset of this 95+ age group were assigned to the healthy subsample in order to provide detailed information on those persons who are nondisabled and over the age of 95. The Informal Caregiver survey was conducted with this wave.

In each of the six surveys, large samples (N=20,000) of the oldest-old population (i.e., those 85 and over) are obtained. The survey data (i.e., detailed community and institutional interviews; screening surveys of nondisabled persons) were linked to Medicare service use files for the years 1982 to 2004. The linkage to Medicare enrollment files between 1982 and 2004 was 100 percent, i.e., there was complete follow-up of all cases (including survey nonrespondents) for Medicare eligibility (and for most years, detailed Part A and B use), mortality, and date of death. Medicare mortality records (and dates of death) are available for 1982 to 2005.

In the new 1982, 1984, 1989, 1994, 1999, and 2004 NLTC file linked to Medicare service use and cost data 1982 to 2004, a number of ancillary analytic variables and special longitudinal weights were appended to aid the investigator in dealing with the complex cross temporal studies of the survey. This will become increasingly important as, with the 1999 file, meaningful cohort analyses can be conducted. Medicare benefit changes can be correlated with period changes in disability and health. The number of deaths (i.e., about 32,000 from 1982 to 2005) is large enough that detailed mortality analyses can be done. Over the 22 years spanned by the six surveys, a total of 49,242 distinct individuals were followed from and linked to Medicare records. The 1994 survey file was added to the public use linked file during 1996. The 1999 survey's public use file is available for the Center for Demographic Studies. The 2004 survey's public use file is anticipated to be released in 2006.

Data Availability: The Beta-1 version of the 2004 NLTC is available at <http://www.icpsr.umich.edu/NACDA/archive.html>. The 1982, 1984, 1989, 1994, 1999, and 2004 National Long Term Care Survey Public Use Files are available through ICPSR. Continuously linked Medicare data (1982 thru 2004) for the persons in the 1982, 1984, 1989, 1994, and 1999 National Long Term Care Surveys have been placed on CD-ROM. These data are only available from CMS. CMS contact information is provided below. All records have a unique sequence number to link to the National Long Term Care Survey respondents. Documentation for all years has been included on the CD for ease of use.

For more information:

Contact: Georganne Patmios, National Institute on Aging

Email: [patmiosg@nia.nih.gov](mailto:patmiosg@nia.nih.gov)

Phone: 301-496-3138

## **New Immigrant Survey (NIS)**

Sponsors: National Institute on Aging, National Institute of Child Health and Human Development, Office of Behavioral and Social Science Research, National Science Foundation, U.S. Citizenship and Immigration Services, Assistant Secretary for Planning and Evaluation in HHS, Pew Charitable Trusts.

Data Collection Agency/Organization: NORC (first round).

Purpose: Immigration has a major impact on American society—it accounts for one-third of U.S. population growth, and the United States takes in more immigrants than all other countries combined. In perhaps no other area of public policy is there such a large gap between information needs and existing data. Immigration processes and immigration policy continue to be the subject of much political and scientific debate. What are the contributions and costs of immigrants to the economy? What is the relationship between legal and illegal immigration? What is known about the origins of legal immigrants, how many return to their home countries, and what factors affect their assimilation?

The main objective is to provide a public use database on new legal immigrants to the United States that can address scientific and policy questions about migration behavior and the impacts of migration. The aims of the study are:

- To assess the differences that occur within immigrant lifestyles pre- and post-immigration, how social networks of families serve as support, and family reunification
- To evaluate employment situations prior to immigration, i.e., occupation, income, and social assimilation
- To evaluate the assimilation of immigrants into U.S. society
- To examine the schooling of immigrants
- To compare NIS survey instruments with comparable instruments used in major U.S. longitudinal surveys
- To examine the transition from temporary to permanent citizenship
- To compare the health and wellbeing of immigrants with native citizens

Description: The NIS is a multicohort prospective-retrospective panel of new legal immigrants to the United States. It was designed following specific recommendations from the National Research Council and other scientific panels, which have long advocated the creation of a longitudinal study of immigrants to overcome the limitations of government statistics and extant databases with information on the foreign-born. A survey pilot project, the NIS-P, was carried out in 1996 to inform the fielding and design of the full NIS. Baseline interviews were ultimately conducted with 1,127 adult immigrants. Sample members were interviewed at baseline, 6 months and 12 months, with half of the sample also interviewed at 3 months.

The first full cohort, NIS-2003, is based on a nationally representative sample of the electronic administrative records compiled for new immigrants by the U.S. government. NIS-2003 sampled immigrants in the period May–November 2003. The geographic sampling design takes advantage of the natural clustering of immigrants. It includes all top 85 Metropolitan Statistical

Areas (MSAs) and all top 38 counties, plus a random sample of other MSAs and counties. Interviews were conducted in respondents' preferred languages. The baseline was multimodal: 60 percent of adult interviews were administered by telephone; 40 percent were in-person. The baseline round was in the field from June 2003 to June 2004, and includes in the Adult Sample 8,573 respondents, 4,336 spouses, and 1,072 children aged 8–12.

**Questionnaire Topics:** Several modules of the NIS were designed to replicate sections of the continuing surveys of the U.S. population that provide a natural comparison group. Other modules and questions are unique to the NIS.

**Health:** Self-reports of conditions, symptoms, functional status, smoking, and drinking history, use/source/costs of health care services, depression, pain.

**Background:** Childhood history and living conditions, education, migration history, marital history, military history, fertility history, language skills, employment history in the United States and foreign countries, social networks, religion.

**Family:** Rosters of all children; for each, demographic attributes, education, current work status, migration, marital status and children; for some, summary indicators of childhood and current health, language ability.

**Transfers:** Financial assistance given/received to/from respondent from/to relatives, friends, employer.

**Economic:** Sources and amounts of income, including wages, pensions, and government subsidies; type, value of assets and debts.

**Housing environment:** Type, ownership of consumer durables.

**Data Availability:** Public-release data from the NIS-2003 baseline and the NIS-P are housed at the Princeton Office of Population Research Data Archive, accessible via the project website: <http://nis.princeton.edu/>. Users must complete a short registration process the first time they access the data.

For more information:

Contact: Jennifer Martin, Project Manager, Princeton University

Email: [nis@opr.princeton.edu](mailto:nis@opr.princeton.edu)

Website: <http://nis.princeton.edu>



## **Panel Study of Income Dynamics (PSID)**

Sponsors: National Institute on Aging, National Institute of Child Health and Human Development, National Science Foundation, Assistant Secretary for Planning and Evaluation in HHS, U.S. Department of Agriculture.

Data Collection Agency/Organization: University of Michigan, Institute for Social Research

Purpose: The Panel Study of Income Dynamics, begun in 1968, provides a wide variety of information about families and individuals, as well as some information about the locations in which sample families reside. The central foci of the data are economic and demographic, with substantial detail on income sources and amounts, employment, family composition changes, childbirth and marriage histories, and residential location. Content of a more sociological or psychological nature is also included in some waves of the study.

Description: Over the life of the PSID, the National Institute on Aging has funded supplements on wealth, health, parental health and long term care, and the financial impact of illness. The NIA has also funded health and housing supplements to the PSID database in order to identify and predict situations of dependent care among individuals aged 55 and over and to model retirement and residential mobility. Global health questions were asked in years prior to 1999. Starting in 1999, much greater detail on specific health conditions and health care expenses is included. In 2001, a question series about 30-day emotional distress was added. In 2003, the two stem questions from the Composite International Diagnostic Interview were added to assess symptoms of 12-month major depression. PSID household wealth measures are included in the 1984, 1989, 1994 and 1999–2003 waves. Savings measures are incorporated in 1989, 1994 and 1999–2003. The 1999 wave contains extensive questions on pensions of the Head and Wife, and a special NIA-sponsored data collection of Head's and Wife's employer pension plans is currently being analyzed. The 2001–03 waves added a supplement on philanthropic giving and volunteering. A question series on internet and computer use was added in 2003.

In 1996, the PSID began a special compilation of all persons ever in PSID families who were known to have died—information for more than 4,000 individuals through the 1997 wave, including death dates. Cause of death information from the National Death Index (NDI) will be added to the file. The resulting dataset will be released under confidential contract. For each wave from 1999 on, the PSID will update this file with newly discovered deaths and match these new cases to the NDI for cause of death. Because of its panel nature, the PSID is unique in having such a sizeable number of deceased individuals, yielding adequate power for analysis of the relative contributions of various factors to mortality hazards. This project will further the understanding of mortality and morbidity processes by providing dates and causes of death information in a well-established long-term longitudinal dataset that also contains information on generational links and socioeconomic and health conditions of individuals over time. These data will allow researchers to analyze the causal linkages among these measures. Moreover, with the availability of environmental measures in the PSID, such as the social and family history variables and GIS-linked environmental data, researchers will be able to investigate the impact of environmental factors on long-term health and eventual mortality. These kinds of studies will

inform the persistent question of the relative power of different factors shaping health and mortality.

The PSID has collected some basic health information over most waves. The measures apply to both the Head and Wife, irrespective of age. Coverage of the full age range, which has been in place since the 1999 wave, allows a life course perspective for researchers. The measures include health limitations in work and everyday life. In 2001, a question series on 30-day nonspecific emotional distress was added. Questions assessing the symptoms of major depression over the past 12 months were included in 2003. Available online is a report assessing the quality of health data in the PSID and providing comparative analyses with NHIS, “Analysis of the Quality of the Health Data in the PSID”:

[http://psidonline.isr.umich.edu/Guide/Quality/report\\_on\\_health\\_qsv2.pdf](http://psidonline.isr.umich.edu/Guide/Quality/report_on_health_qsv2.pdf). This assessment revealed a high degree of comparability between PSID and NHIS in the prevalence estimates of a range of health conditions and health behaviors.

A planned separate release concerns data on pension plans. Basic pension information and the name of the employer of the Head and Wife were collected in the 1999 wave, and the named employers were contacted in 2000 for copies of Summary Plan Descriptions (SPD) of employee pensions. From this the PSID will use the individual characteristics of Head and Wife to impute pension wealth under different retirement assumptions and include these estimates on a public release Internet file. Because of the long earnings histories available in the PSID, estimates of Social Security wealth could be constructed without resorting to confidential Social Security earnings records.

**Study Design:** Reporting unit is the family: single person living alone or sharing a household with other nonrelatives; group of people related by blood, marriage, or adoption; unmarried couple living together in what appears to be a fairly permanent arrangement. Respondent is usually the family Head, who is usually the major adult male earner. Interviews conducted annually from 1968 through 1997. Biennial interviewing began in 1999. Event history calendar methodology added in 2001 to facilitate recall of employment spells. Oversample of black persons (30 percent). Waves 1990 through 1995 included a 20 percent Hispanic oversample; within the Hispanic oversample, Cubans and Puerto Ricans were oversampled relative to Mexicans.

**Data Availability:** Main data files, that is, cross-year individual files and annual family files, are updated with each subsequent wave of data. Other special public-release files include the 1968–80 Retrospective Occupation-Industry Files; the 1985 Ego-Alter File; the 1968–85 Relationship File; the 1988 Time and Money Transfers File; the 1985–2001 Childbirth and Adoption History File; the 1985–2001 Marriage History File; the 2001 Parent Identification File; two 1990 Health Supplement Files; the 1991 Parent Health Supplement file; the 1993 Health Care Burden File; the 1984, 1989, 1994, 1999, and 2001 Wealth Supplement Files, which provide details on the level of various types of assets; and the 1994–2001 Hours of Work and Wage Files.

Restricted files include the Geocode Match File with information for 1968 through 2001, the 1968–2001 Death File, and the 1991 Medicare Claims File.

The Internet provides access to main release and archive files. All data from 1994 through 2001 are available as public release files; prior waves can be obtained in archive versions. No data were collected in 1998 or 2000 or 2002, as the study was moved to a 2-year period. The special files with weights for families are also available on the website.

For more information:

Contact: Frank Stafford, University of Michigan

Email: [fstaffor@isr.umich.edu](mailto:fstaffor@isr.umich.edu)

Phone: 734-763-5166

Website: <http://psidonline.isr.umich.edu>

## **Wisconsin Longitudinal Study (WLS)**

Sponsors: National Institute on Aging

Data Collection Agency/Organization: University of Wisconsin

**Purpose:** The Wisconsin Longitudinal Study (WLS) is a long-term study based on a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. A companion sample contains comparable data for a randomly selected sibling of most respondents. WLS data cover social background, youthful aspirations, schooling, military service, labor market experiences, family characteristics and events, social participation, psychological characteristics, health and well-being, and retirement.

The WLS provides an opportunity to study the life course, intergenerational transfers and relationships, family functioning, physical and mental health and well-being, and morbidity and mortality from late adolescence through middle age. The WLS is a long-term study of a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957 and of their randomly selected brothers and sisters. Survey data were collected from the original respondents or their parents in 1957, 1964, 1975, and 1992, and from a selected sibling in 1977 and 1993. These data provide a full record of social background, youthful aspirations, schooling, military service, family formation, labor market experiences, and social participation of the original respondents. The survey data from earlier years have been supplemented by mental ability tests (of primary respondents and 2000 of their siblings), measures of school performance, and characteristics of communities of residence, schools and colleges, employers, and industries. The WLS records for primary respondents are also linked to those of three same-sex high school friends within the study population. In 1977 the study design was expanded with the collection of parallel interview data for a highly stratified sub-sample of 2000 siblings of the primary respondents. In the 1992-93 round of the WLS the sample was expanded to include a randomly selected sibling of every respondent with at least one brother or sister, and the content was extended to obtain detailed occupational histories and job characteristics; incomes, assets, and inter-household transfers; social and economic characteristics of parents, siblings, and children and descriptions of the respondents' relationships with them; and extensive information about mental and physical health and well-being.

The WLS sample is broadly representative of white, non-Hispanic American men and women who have completed at least a high school education. Among Americans aged 50 to 54 in 1990 and 1991, approximately 66 percent were non-Hispanic white persons who completed at least 12 years of schooling. Some strata of American society are not well represented. The WLS sample is mainly of German, English, Irish, Scandinavian, Polish, or Czech ancestry. It is estimated that about 75 percent of Wisconsin youth graduated from high school in the late 1950s – everyone in the primary WLS sample graduated from high school; about seven percent of their siblings did not graduate from high school. Minorities are not well-represented: there are only a handful of African American, Hispanic, or Asian persons in the sample, though a project is currently under way to find all African Americans who graduated from Wisconsin high schools in 1957. About 19 percent of the WLS sample is of farm origin, and that is consistent with national estimates of

persons of farm origin in cohorts born in the late 1930s. As in the later, large, longitudinal studies of school-based samples, age variation occurs in repeated observations rather than in cross-section. Also, siblings cover several adjoining cohorts: they were born primarily between 1930 and 1948. In 1964, 1975, and again in 1992, about two-thirds of the sample lived in Wisconsin, and about one-third lived elsewhere in the U.S. or abroad.

In 1992-93 a follow-up survey – of about 9,000 men and women who were first interviewed as seniors in Wisconsin high schools in 1957 and have subsequently been followed up in 1957, 1964, and 1975 – was completed. Most respondents were 53 or 54 years old when interviewed. Other members of the original sample who were not interviewed in 1975 (475 of 850 surviving non-respondents) have also been interviewed. In all, 8493 of the 9741 surviving members of the original sample have been interviewed. In 1993-94 randomly selected siblings of the high school graduates were interviewed. Some 2000 siblings were previously interviewed in 1977, and they and approximately 2800 additional siblings were interviewed in the 1993-94 round of the study. The surveys included a one-hour telephone interview followed by a 20-page, self-administered questionnaire. Brief, close-out interviews were carried out with a relative of respondents who have died, and, in cases in which the selected sibling has died, close-out data from the original respondent were obtained.

Data availability: Public Use Files are currently available from all collection rounds except the widows, which will be available in the future. Please visit <http://www.ssc.wisc.edu/wlsresearch/> for more information.

For more information:

Contact: Wisconsin Longitudinal Study

Department of Sociology

1180 Observatory Drive

Madison, Wisconsin 53706

Email: [wls@ssc.wisc.edu](mailto:wls@ssc.wisc.edu)

Fax: 608-262-8400

## **Wisconsin National Survey of Midlife Development in the United States: MIDUS 1 (1995-1996) and MICUS 2 (2004-2005)**

Sponsors: National Institute on Aging and The John D. and Catherine T. MacArthur Foundation

Data Collection Agency/Organization: University of Wisconsin

Purpose: MIDUS (Midlife in the U.S.) is a national study of the role of behavioral, psychological, and social factors in health and well-being, broadly defined, as people age from early adulthood into midlife and old age. It was conceived by a multidisciplinary team of scholars from fields of psychology, sociology, epidemiology, demography, anthropology, medicine, and health care policy. The first wave of data collection utilized survey and questionnaire methodologies. The second wave of the study has incorporated numerous additions, including new measures of cognitive performance, diverse biomarkers, and neurological assessments.

Description: In 1995-96, the MacArthur Midlife Research Network carried out a national survey (MIDUS I) of over 7,000 Americans aged 25 to 74. Wave 1 respondents were drawn from a nationally representative random-digit-dialing sample of non-institutionalized, English-speaking adults. Also included was a large sample of siblings of the main respondents as well as a national sample of twins, representing the same age ranges as the RDD sample.

Main Sample: Contains responses from the main survey of 4,242 respondents. Respondents were asked to provide extensive information on their physical and mental health throughout their adult lives, and to assess the ways in which their lifestyles, including relationships and work-related demands, contributed to the conditions experienced. Those queried were asked to describe their histories of physical ailments, including heart-related conditions and cancer, as well as the treatment and/or lifestyle changes they went through as a result. A series of questions addressed alcohol, tobacco, and illegal drug use, and focused on history of use, regularity of use, attempts to quit, and how the use of those substances affected respondents' physical and mental well-being. Additional questions addressed respondents' sense of control over their health, their awareness of changes in their medical conditions, commitment to regular exercise and a healthy diet, experience with menopause, the decision-making process used to deal with health concerns, experiences with nontraditional remedies or therapies, and history of attending support groups. Respondents were asked to compare their overall well-being with that of their peers and to describe social, physical, and emotional characteristics typical of adults in their 20s, 40s, and 60s. Information on the work histories of respondents and their significant others was also elicited, with items covering the nature of their occupations, work-related physical and emotional demands, and how their personal health had correlated to their jobs. An additional series of questions focusing on childhood queried respondents regarding the presence/absence of their parents, religion, rules/punishments, love/affection, physical/verbal abuse, and the quality of their relationships with their parents and siblings.

Siblings and Twins: The Survey also contains similar data from a survey of 951 respondents, each of whom was a sibling of a respondent in the main survey. Siblings participated in the same assessments as the respondents. The Survey also contains twins data, from a separate

national survey unrelated to the main MIDUS survey. From this separate national survey, a total of 1,996 twins agreed to participate. The Twins respondents were given the same assessments as the Main and Siblings samples. Additionally, the Twins sample was asked a series of questions about their birth, shared physical characteristics, childhood and adult relationships with their twin, whether they were dressed alike as children, and whether others experienced difficulty identifying them correctly.

In 2002, the National Institute on Aging awarded funds to the Institute on Aging at the University of Wisconsin-Madison to conduct a longitudinal follow-up of all original participants. Called MIDUS II, this second wave of the study includes five research projects, covering the following topics:

Project 1 - provides longitudinal follow-up on the psychosocial, sociodemographic, and health variables assessed in MIDUS I; recruits a new city-sample (Milwaukee, WI) of African Americans.

Project 2 - provides longitudinal follow-up on the daily stresses on a subsample of MIDUS I respondents, includes daily measures of salivary cortisol on these individuals.

Project 3 - includes new cognitive assessments (including speed of processing) for the full MIDUS sample, plus longitudinal follow-up for the Boston cognitive subsample from MIDUS I.

Project 4 - includes comprehensive biomarker assessments (neuroendocrine, immune, cardiovascular, musculoskeletal) as laboratory challenge studies on a subsample of MIDUS respondents, collected at one of three General Clinical Research Centers (Georgetown University; University of California-Los Angeles, University of Wisconsin-Madison, Georgetown).

Project 5 - includes brain assessments (EEG, neuroimaging) on a subsample of respondents, including minorities, in the biomarker study (Project 4).

Data Availability: Data and comprehensive documentation for MIDUS I and II are available via ICPSR (Inter-University Consortium for Political and Social Research) website:

<http://www.icpsr.umich.edu>.

For more information:

Contact:

MIDUS - A National Study of Health and Well-Being

University of Wisconsin - Madison Institute on Aging

1300 University Avenue, 2245 MSC

Madison, Wisconsin 53706

Phone: 608-262-2056

Fax: 608-263-6211

Website: <http://www.midus.wisc.edu>

Office of the Assistant Secretary for Planning and Evaluation

**National Study of Assisted Living for the Frail Elderly, 2000**

Sponsor: Office of the Assistant Secretary for Planning and Evaluation in HHS

Data Collection Agency/Organization: Research Triangle Institute

Purpose: Hawes et al. (2000) fielded the first national survey in 1998 that focused exclusively on one component of residential care—assisted living. To be included in the survey, a facility had to have more than 10 beds, largely serve an elderly population, and either describe itself as an assisted living facility or offer services that included 24-hour staff oversight, housekeeping, at least two meals a day and personal assistance (defined as assistance with at least two of the following: medication management, bathing or dressing). Hawes et al. excluded smaller places to avoid collecting information on smaller board and care homes; therefore it is likely that this approach missed residential care places serving the elderly that might meet a broader definition of assisted living.

A primary goal of the study was to describe facility and resident characteristics, and categorize facilities by level of service and degree of privacy. One of the main findings from the study was that acuity levels were the highest, and admission and retention policies the most expansive, in facilities categorized as “high service/low privacy.” Hawes et al. also found that there was significant variability in the assisted living industry and a substantial portion of the facilities did not include components that were consistent with the assisted living philosophy, i.e., an emphasis on privacy and a homelike environment. The study also found that assisted living in the early and mid-1990s was generally not affordable to low- and moderate-income persons.

Reports: See <http://aspe.hhs.gov/daltcp/reports/facres.htm>

For more information:

Contact: Gavin Kennedy

Email: [Gavin.Kennedy@hhs.gov](mailto:Gavin.Kennedy@hhs.gov)

Phone: 202-690-6443



## **Substance Abuse and Mental Health Services Administration**

### **Client/Patient Sample Survey (CPSS)**

Sponsor: Substance Abuse and Mental Health Services Administration, Center for Mental Health Services

Data Collection Agency/Organization: Center for Mental Health Services

Purpose: To provide national estimates on the number and sociodemographic, clinical, and service use characteristics of persons who receive services in specialty mental health organizations nationwide.

Survey Universe: The specialty mental health organizations include: State/county mental hospitals, private psychiatric hospitals, the separate psychiatric services of non-Federal general hospitals and Veterans Administration medical centers, multiservice mental health organizations, residential treatment centers for emotionally disturbed children, and freestanding outpatient clinics and partial care organizations.

Research Design: The samples are based on a two-staged, stratified design. For each type of program (e.g., inpatient, outpatient), clients/patients are systematically sampled from within a sample of specialty mental health organization types. The sample cases are weighted to produce national-level estimates on the number of persons served.

Survey Mode: Mailed client/patient questionnaires to programs for completion on a sample of program's caseload; data abstracted from medical records.

Unit of Analysis: Cross-tabulations of client/patient-level characteristics by type of program and type of specialty mental health organization.

Sample: Approximately 2,500 programs; approximately 23,000 clients/patients.

Topics/Key Words: Inpatient mental health care, residential mental health care, outpatient mental health care, age, gender, race, Hispanic origin, marital status, veteran status, legal status, living arrangement, type of residence, presenting problem, psychiatric diagnosis, dual diagnosis, medical condition, prior mental health care, referral source, referral on discharge, principal payment source, types of services received, level of functioning.

Data Availability: 1970, 1975, 1980, 1986, 1997.

Linked Data: The National Survey of Mental Health Treatment Facilities (NSMHTF), formerly The Survey of Mental Health Organizations (SMHO).

Data Dissemination: Public-use data files available upon request.

Reports: *Mental Health, United States, 2004*; *Mental Health, United States, 2002*; *Mental Health United States, 2000*; *Mental Health, United States, 1992*; *Mental Health United States, 1987*; *Mental Health, United States, 1985*.

Future Plans: A CPSS was conducted in late spring 2007. The less than 24-hour care programs (e.g., outpatient) of specialty mental health organizations were surveyed. The 2007 CPSS also included an adult self-report consumer perception of care survey component. Program respondents (and adult consumer respondents) had the option of completing questionnaires on paper or electronically. A public-use data file will be available upon request winter 2009.

For more information:

Contact: Laura Milazzo-Sayre

Email: [laura.milazzo-sayre@samhsa.hhs.gov](mailto:laura.milazzo-sayre@samhsa.hhs.gov)

## **Drug Abuse Warning Network (DAWN)**

Sponsor: Substance Abuse and Mental Health Services Administration, Office of Applied Studies

Data Collection Agency: Substance Abuse and Mental Health Services Administration

Purpose: The Drug Abuse Warning Network (DAWN) is an ongoing, national public health surveillance system that collects data and reports information on adverse health consequences associated with drug use, misuse, and abuse. Specifically, DAWN captures data on drug-related ED visits from a national probability sample of hospitals with oversampling in selected metropolitan areas. DAWN also collects data on drug-related deaths reviewed by medical examiners and coroners (ME/Cs) in participating jurisdictions in selected metropolitan areas and States. DAWN helps communities, member facilities, local public health authorities, as well as Federal agencies and policymakers assess emerging drug problems, improve patient care, and monitor drug trends. The entire DAWN system was recently redesigned to collect better data more efficiently and to serve the information needs of its users more effectively.

Survey Universe: Hospitals are selected by scientific sampling. Hospitals eligible for the sample are short-term, general, non-Federal facilities, with 24-hour emergency departments. The sample is stratified based on these hospital characteristics:

- Location (in selected metropolitan areas vs. elsewhere in the United States)
- Ownership (public vs. private)
- Size (small, medium, or large)

ME/Cs are invited to join DAWN based on their location in selected metropolitan areas and States across the country. DAWN invites jurisdictions:

- In the same metropolitan areas where DAWN has oversampled hospitals. With ME/C jurisdictions and hospital emergency departments covering the same metropolitan areas, so that DAWN can measure drug-related morbidity and mortality for the same population base.
- In States with centralized medical examiner systems.
- In selected metropolitan areas where DAWN does not have an active sample of hospitals. In some instances, this will be the only data on substance abuse readily available for these local areas.

Survey Mode: Drug-related emergency department visits are found through a process of retrospective chart review. All charts for patients treated in a participating emergency department are reviewed to find the drug-related visits. Data items are submitted electronically for the drug-related visits found. A comparable process is followed in medical examiner/coroner jurisdictions where all case files are reviewed to find drug-related deaths. Data items are submitted electronically on the drug-related deaths.

Unit of Analysis: The drug-related ED visit or the drug-related death is the primary unit of analysis used in DAWN. A secondary unit of analysis is the drug report. Multiple drugs may be reported for an individual visit or death.

Sample: The current DAWN sample design consists of a stratified, single-stage cluster sample. The design uses primary geographic strata with as many as eight finer strata within each geographic stratum. The geographic strata represent the selected metropolitan areas and subdivisions and the balance of the United States (“remainder areas”). The stratification within each geographic area reflects the public or private ownership of the hospital and the size of the hospital in terms of the number of emergency department visits. Hospitals are selected with equal probability within stratum and all emergency department visits, charts reviewed, and DAWN cases are captured within sampled hospitals. In 2006, 205 hospitals submitted data that were used for estimation.

Topics: Drug-related emergency department (ED) visits and drug-related deaths investigated by medical examiners and coroners (ME/Cs).

DAWN includes all types of drugs:

- Illegal drugs of abuse
- Prescription and over-the-counter medications
- Dietary supplements
- Nonpharmaceutical inhalants
- Alcohol in combination with other drugs (adults and children)
- Alcohol alone (age < 21).

Data Availability: DAWN makes available numerous publications, tables, estimates, and custom tabulations periodically and in response to special requests. DAWN collects individually identifiable data, which cannot be disclosed without consent. Therefore, there is no public use data file from DAWN.

Linked Data: N/A

Reports: DAWN publishes annual estimates of drug-related emergency department visits for the Nation and for selected metropolitan oversample areas. Estimates for 2004 were the first to reflect the major redesign of the DAWN system that began in 2003. Therefore, no comparisons of DAWN estimates prior to 2004 are possible. DAWN publishes annual profiles of drug-related deaths for selected metropolitan areas, jurisdictions, and States. Mortality profiles for 2003 were the first to reflect all the changes from the redesign. Therefore, no comparisons of DAWN mortality data prior to 2003 are possible. The DAWN Report is a series of short publications on special topics that are published periodically from DAWN. Methodology publications are produced periodically to address the statistical design, analyses, and approaches used in DAWN.

For more information:

Contact: Kathy Poneleit

Email: [Kathy.Poneleit@samhsa.hhs.gov](mailto:Kathy.Poneleit@samhsa.hhs.gov)

Phone: 240-276-1254

Website: <http://dawninfo.samhsa.gov/default.asp>

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

Sponsor: Substance Abuse and Mental Health Services Administration, Office of Applied Studies

Data Collection Agency/Organization: Office of Applied Studies

Purpose: The National Survey on Drug Use and Health (NSDUH) has been conducted since 1971 and serves as the primary source of information on the prevalence and incidence of illicit drug, alcohol, and tobacco use in the civilian, noninstitutionalized population aged 12 or over in the United States. Information about substance abuse and dependence, mental health problems, and receipt of substance abuse and mental health treatment also is included. Before 2002, the name of the survey was the National Household Survey on Drug Abuse (NHSDA).

Survey Universe: The civilian, noninstitutionalized population aged 12 years old or over residing within the United States. The survey covers residents of households (living in houses/townhouses, apartments, condominiums, etc.), persons in noninstitutional group quarters (e.g., shelters, rooming/boarding houses, college dormitories, migratory workers' camps, halfway houses), and civilians living on military bases. Persons excluded from the survey include homeless people who do not use shelters, active military personnel, and residents of institutional group quarters, such as correctional facilities, nursing homes, and long-term psychiatric hospitals.

Research Design: Ongoing cross-sectional face-to-face survey.

Survey Mode: The data are collected through in-person interviews with sample persons, incorporating procedures that would be likely to increase respondents' cooperation and willingness to report honestly about their illicit drug use behavior. Confidentiality is stressed in all written and oral communications with potential respondents. Respondents' names are not collected with the data, and CAI (computer-assisted interviewing) methods, including audio computer-assisted self-interviewing (ACASI), are used to provide a private and confidential setting to complete the interview.

Sample: A 50-State design has been employed during 2004-2009, where 8 States were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) with samples large enough to support direct State estimates. Sample sizes in these States are around 3,600. For the remaining 42 States and the District of Columbia, sample size is around 900.

States were first stratified into a total of 900 sub-state sampling (SS) regions (48 regions in each large sample State and 12 regions in each small sample State). These regions were contiguous geographic areas designed to yield the same number of interviews on average. In each SS region, 8 census tracts were selected with probability proportional to size for inclusion in the survey each year. Within these tracts, adjacent census blocks were combined to form the

second-stage sampling units, called area segments. One segment per census tract per year was selected with probability proportional to population size.

These sampled segments were allocated equally into four separate samples, one sample (consisting of 2 segments) for each 3-month period (calendar quarter) during the year, so that the survey is essentially continuous in the field. In each of these area segments, a listing of all addresses is made, from which a sample of about 170,000 addresses are selected each year. Of the selected addresses, about 140,000 are eligible sample units. In these sample units (which can be either households or units within group quarters), sample persons are randomly selected using an automated screening procedure programmed in a handheld computer carried by the interviewers. The number of sample units completing the screening is about 130,000. Youths aged 12–17 years and young adults aged 18–25 years are oversampled at this stage. Because of the large sample size, there no longer is a need to oversample racial/ethnic groups, as was done on surveys prior to 1999. A total of about 82,000 persons are selected nationwide leading to an annual final respondent sample of about 67,500 persons. This sample is representative of the U.S. general population (since 1991, the civilian, noninstitutionalized population) aged 12 and over.

Topics: A number of key measures of substance use and mental health are reported from the NSDUH data. For illicit drug use, alcohol use, and tobacco use, information is presented about use in the lifetime, past year, and past month. Use in the past month also is referred to as “current use.” The survey also produces measures of abuse, dependence, treatment, and mental health problems, generally for the past year.

Measures for nine specific classes of drugs are presented in analyses. These include the use of marijuana/hashish, cocaine (including crack), inhalants, hallucinogens, heroin, and prescription-type drugs used nonmedically (pain relievers, tranquilizers, stimulants, and sedatives). Nonprescription medications and legitimate uses under a doctor’s supervision are not included in the survey. Summary measures such as “any illicit drug use” and “any illicit drug use other than marijuana” are produced.

In addition to a measure of any alcohol use, measures of “binge” alcohol use and heavy use in the past 30 days have been developed. “Binge” alcohol use is defined as having five or more drinks on the same occasion at least once in the 30 days prior to the survey, and “heavy” use is defined as having five or more drinks on the same occasion on at least 5 different days in the past 30 days.

The measure of tobacco use includes use of cigarettes, chewing tobacco, snuff, cigars, and pipe tobacco.

Measures of the perceived risk of harm from use of a number of illicit drugs, alcohol, and cigarettes at varying levels of use also are included. Measures of substance abuse and dependence and depression are based on criteria described in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM–IV).

In addition to these key substantive terms, a number of demographic and geographic characteristics are used in many analyses of the NSDUH data.

Data Availability: Public-use data files for 1979, 1982, 1985, 1988, and annually from 1990 to 2007 are currently available through the Substance Abuse and Mental Health Data Archive (SAMHDA) and the archive's on-line data analysis system (<http://www.icpsr.umich.edu/SAMHDA/>).

Linked Data: N/A

Reports: A complete listing of previously published NSDUH reports is available from SAMHSA's Office of Applied Studies. Many of these reports are available on the SAMHSA website.

The NSDUH Report, published approximately twice a month, presents key findings from the NSDUH. These reports are available by mail and are posted on the SAMHSA website.

Associated with the release of the annual survey findings are detailed tables presenting analyses of substance use and other measures by demographic and geographic characteristics. These are also posted on the SAMHSA website.

For More Information:

Website: <http://oas.samhsa.gov/nsduh.htm>



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

Sponsor: Substance Abuse and Mental Health Services Administration, Office of Applied Studies

Data Collection Agency/Organization: Office of Applied Studies

Purpose: N-SSATS is designed to collect information from all facilities in the United States, both public and private, that provide substance abuse treatment. N-SSATS provides the mechanism for quantifying the dynamic character and composition of the U.S. substance abuse treatment delivery system. The objectives of N-SSATS are to collect multipurpose data that can be used to:

- Assist SAMHSA and State and local governments in assessing the nature and extent of services provided in State-supported and other treatment facilities and in forecasting treatment resource requirements.
- Update SAMHSA's Inventory of Substance Abuse Treatment Services (I-SATS), which includes all known drug and alcohol abuse treatment facilities.
- Analyze treatment services trends and conduct comparative analyses for the nation, regions, and States.
- Generate the National Directory of Drug and Alcohol Abuse Treatment Programs, a compendium of facilities approved by State substance abuse agencies for the provision of substance abuse treatment.
- Update the information in SAMHSA's Substance Abuse Treatment Facility Locator, a searchable database of facilities approved by State substance abuse agencies for the provision of substance abuse treatment. The Treatment Facility Locator is available on the Internet at <http://findtreatment.samhsa.gov/>.

Survey Universe: Alcohol and drug treatment facilities and services throughout the 50 States, the District of Columbia, and other U.S. jurisdictions, including facilities that are licensed, certified, or otherwise approved by the State substance abuse agency to provide substance abuse treatment. A second group of facilities including treatment facilities that State substance abuse agencies, for a variety of reasons, do not fund, license, or certify. Many of these facilities are private for-profit, small group practices, or hospital-based programs. Most of them are identified through periodic screening of alternative source databases. This group also includes programs operated by Federal agencies, the Department of Veterans Affairs (VA), the Department of Defense, and the Indian Health Service. The 2007 N-SSATS facility universe included 1 16,773 active treatment facilities.

Research Design: N/A

Survey Mode: The N-SSATS is a mixed-mode survey. All facilities are sent a paper questionnaire but are also given the option of completing a web-based questionnaire. Facilities that do not complete and return a paper questionnaire or web-based questionnaire are contacted by telephone and are asked to complete a telephone interview.

Unit of Analysis: N-SSATS is designed to collect data from each physical location where treatment services are provided. Accordingly, SAMHSA requests that State substance abuse agencies use the point of delivery of service (i.e., physical location) as the defining factor for a facility. Because of the different State administrative systems, however, there are some inconsistencies in implementation. For example, in some States, multiple treatment programs (e.g., detoxification, residential, and outpatient) at the same address and under the same management have separate State licenses. These are treated as separate by the State substance abuse agency, and are given separate I-SATS ID numbers. In other States, multiple sites are included as a single entity under a parent or administrative unit. In many of these cases, individual sites can report services data in N-SSATS, but client data are available only at a higher administrative level.

Sample: See Survey Universe above.

Topics: Topics include facility identification information; facility focus (substance abuse treatment services, mental health services, mix of mental health and substance abuse treatment services, general health care, other); facility operation (private for-profit; private non-profit; government—local, county, or community; State; Federal; tribal); hotline operation; services offered (assessment services, substance abuse therapy and counseling, pharmacotherapies, testing, transitional services, other services); operation of an Opioid Treatment Program (OTP) certified by the Substance Abuse and Mental Health Services Administration; services in sign language for the hearing impaired; services in languages other than English; types of clients accepted into treatment (adolescents, clients with co-occurring mental and substance abuse disorders, criminal justice clients, persons with HIV or AIDS, gays or lesbians, seniors or older adults, adult women, pregnant or postpartum women, adult men, DUI/DWI offenders); special programs or groups offered for specific types of clients (adolescents, clients with co-occurring mental and substance abuse disorders, criminal justice clients, persons with HIV or AIDS, gays or lesbians, seniors or older adults, adult women, pregnant or postpartum women, adult men, DUI/DWI offenders, other); types of treatment provided; hospital inpatient (detoxification, treatment); nonhospital residential (detoxification; short-term, i.e., 30 days or less; long-term, i.e., more than 30 days); outpatient (detoxification, methadone maintenance, day treatment or partial hospitalization, intensive, regular); payment options; receipt of public funding for substance abuse treatment programs; managed care participation; number of clients in treatment on March 30, 2007 (total, clients under age 18, clients receiving methadone or buprenorphine); number of beds designated for nonhospital residential and hospital inpatient substance abuse treatment; client substance abuse problem treated; facility licensure, certification, or accreditation; and facility website availability.

Data Availability: Public-use data files for 1997-2007 are currently available through the Substance Abuse and Mental Health Data Archive (SAMHDA) and the archive's on-line data analysis system (<http://www.icpsr.umich.edu/SAMHDA/>).

Linked Data: N/A

Reports: A complete listing of previously published N-SSATS reports is available from SAMHSA's Office of Applied Studies. Many of these reports are available on the SAMHSA website.

The DASIS Report, published approximately twice a month, presents key findings from the TEDS and N-SSATS. These reports are available by mail and are posted on the SAMHSA website.

For more information:

Website: <http://oas.samhsa.gov/dasis.htm#nssats3>

**National Survey of Mental Health Treatment Facilities (NSMHTF), formerly The Survey of Mental Health Organizations (SMHO)**

Sponsor: Substance Abuse and Mental Health Services Administration, Center for Mental Health Services

Data Collection Agency/Organization: enter for Mental Health Services

Purpose: In the rapidly changing health care services delivery environment, timely data on the availability and utilization of organizations and point-of-contact facilities that provide mental health services is needed for policy formulation. In the past, the Center for Mental Health Services (CMHS) has funded a biennial Survey of Mental Health Organizations (SMHO) to meet information needs. The 2008 NSMHTF expands the scope of the SMHO to address the need to move from an organizational survey to a point-of-contact or facility-based survey to be fielded among the entire inventory of specialty mental health organizations, and in addition, to become more aligned with the National Survey of Substance Abuse Treatment Services (N-SSATS). It is expected that the NSMHTF will continue to occur biennially on the entire inventory of eligible organizations and point-of-contact facilities.

Survey Universe: The NSMHTF includes public and private psychiatric hospitals, general hospitals with separate psychiatric units, residential treatment centers for children and adults, and large public and private multiservice setting organizations and outpatient MH services organizations.

Research Design: Cross-sectional facility survey.

Survey Mode: Mailed questionnaire with Web-based questionnaire completion and Computer-assisted Telephone Interview (CATI) followup.

Unit of Analysis: Approximately 15,000 point-of-contact facilities nationwide representing approximately 4,400 mental health organizations nationwide.

Sample: None

Topics/Key Words: Inpatient psychiatric services, Outpatient psychiatric services.

Data Availability: 1990, 1992, 1994, 1998, 2000, 2002, and 2004

Linked Data: Client-Patient Sample Survey (CPSS).

Data Dissemination: Public-Use files are available upon request.

Reports: *Mental Health, United States, 2004*; *Mental Health, United States, 2002*.

Future Plans: A planned survey will be conducted in 2010.

For more information:

Contact: Laura Milazzo-Sayre, SAMHSA

Email: [laura.milazzo-sayre@samhsa.hhs.gov](mailto:laura.milazzo-sayre@samhsa.hhs.gov)

Phone: 240-276-1764

## **Treatment Episode Data Set (TEDS)**

Sponsor: Substance Abuse and Mental Health Services Administration, Office of Applied Studies

Data Collection Agency/Organization: Substance Abuse and Mental Health Services Administration, Office of Applied Studies

Purpose: TEDS is an annual compilation of data on the demographic characteristics and substance abuse problems of those admitted for substance abuse treatment. The information comes primarily from facilities that receive some public funding. TEDS records represent admissions rather than individuals, as a person may be admitted to treatment more than one time during a calendar year. The TEDS system is comprised of two major components, the Admissions Data Set and the Discharge Data Set.

Survey Universe: TEDS comprises data that are routinely collected by States in monitoring their individual substance abuse treatment systems. In general, facilities reporting TEDS data are those that receive State alcohol and/or drug agency funds (including Federal Block Grant funds) for the provision of substance abuse treatment. However, differences in State systems of licensure, certification, accreditation, and disbursement of public funds affect the scope of facilities included in TEDS. Treatment facilities that are operated by private for-profit agencies, hospitals, and the State correctional system, if not licensed through the State substance abuse agency, may be excluded from TEDS. TEDS does not include data on facilities operated by Federal agencies (the Bureau of Prisons, the Department of Defense, and the Veterans Administration). All 50 States, the District of Columbia, and Puerto Rico participate in TEDS.

Research Design: N/A

Survey Mode: States collect the data from treatment providers either electronically or on paper forms. Selected data items from the individual State data files are converted to a standardized format consistent across States. These standardized data constitute TEDS.

Unit of Analysis: Admission or discharge record.

Sample: See Survey Universe above. TEDS does not involve statistical sampling methods.

Topics: The TEDS admission dataset includes: demographic information; primary, secondary, and tertiary substances and their route of administration, frequency of use, and age at first use; source of referral to treatment; number of prior treatment episodes; service type, including planned use of methadone; and, psychiatric, social, and economic measures reported by some States. The TEDS discharge dataset includes date of discharge; date of last contact; service at discharge; and, reason for discharge. Beginning in 2006, States began to report additional discharge items including substance use at discharge, employment status, and living arrangement.

Data Availability: Public-use data files for 1992–2006 are currently available through the Substance Abuse and Mental Health Data Archive (SAMHDA) and the archive’s on-line data analysis system (<http://www.icpsr.umich.edu/SAMHDA/>).

Reports: A complete listing of previously published TEDS reports is available from SAMHSA’s Office of Applied Studies. Many of these reports are available on the SAMHSA website.

The DASIS Report, published approximately twice a month, presents key findings from the TEDS and N-SSATS. These reports are available by mail and are posted on the SAMHSA website.

The most recent data for each State is available in the form of summary tables on the SAMHSA website at <http://wwwdasis.samhsa.gov/webt/NewMapv1.htm>.

For more information:

Website: <http://wwwdasis.samhsa.gov/dasis2/teds.htm>

## **Department of Housing and Urban Development Office of Policy Development and Research**

### **American Housing Survey (AHS)**

Sponsor: Office of Policy Development and Research, Department of Housing and Urban Development

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To provide data on the characteristics, condition, financing, and costs of housing in the United States and selected metropolitan areas. Information is also provided on the households that occupy the housing units as well as the neighborhood conditions and amenities.

The American Housing Survey (AHS) collects detailed information on the nation's housing and the people that occupy the housing units. For housing characteristics, such data as unit size, housing costs, mortgage financing, equipment and fuels, recent movers, neighborhood quality, and commute to work are collected. For characteristics on the individuals in the household, such information as individual and family income, race/ethnicity, age, education, disabilities, and assisted living are obtained. Data is collected on all types of housing, including single-family homes, townhouses, apartments, condominiums, cooperatives, mobile homes, and vacant housing units. National and metropolitan data are currently collected biennially in odd numbered years. Data for each of 21 selected Metropolitan Areas are collected every 6 years. The national sample covers approximately 55,000 to 60,000 homes. Each metropolitan area sample covers approximately 3,500 to 4,000 homes. The survey is conducted by the Bureau of the Census for the Department of Housing and Urban Development.

The AHS returns to the same housing units year after year to gather data. Therefore, this survey is ideal for analyzing the flow of households through housing.

The AHS is a household survey that asks questions about the quality of housing in the United States. In gathering information, Census Bureau interviewers visit or telephone the household occupying each housing unit in the sample. For unoccupied units, they obtain information from landlords, rental agents, or neighbors.

The AHS conducts a national survey and a metropolitan area survey. Both surveys are conducted during a 3- to 7-month period.

Current Activities: The printed report and public-use data base for the 2007 National and Metropolitan Surveys have been released. Field work will be conducted in the spring and summer of 2009 for the National Survey and for selected metropolitan areas.

Future Plans: Continue collecting National and selected Metropolitan area data biennially in odd numbered years. The sample will be redesigned based on the 2010 Census and we are likely to



begin sampling metropolitan areas within the national sample so as to increase the sample size and number of metropolitan areas surveyed.

Public-Use Data Files: <http://www.huduser.org/datasets/ahs.html>

Reports: Reports are available as PDF documents from <http://www.census.gov/hhes/www/housing/ahs/nationaldata.html> and <http://www.census.gov/hhes/www/housing/ahs/metropolitandata.html>. Hardcopy versions of the reports are available from HUDUSER at <http://www.HUDUSER.ORG/> or the Census Bureau.

For more information:

Contact: American Housing Survey Branch, U.S. Census Bureau

Email: [ahsn@census.gov](mailto:ahsn@census.gov)

Phone: 301-763-3235

Website: <http://www.census.gov/hhes/www/housing/ahs/ahs.html>

Contact: HUD's Division of Housing and Demographic Analysis

Phone: 202-708-1060

Website: <http://www.huduser.org/datasets/ahs.html>

**Department of Labor  
Bureau of Labor Statistics**

**American Time Use Survey (ATUS)**

Sponsor: Bureau of Labor Statistics

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: The American Time Use Survey (ATUS) measures how people living in the United States spend their time. Estimates show the kinds of activities people do and the time spent doing them by sex, age, educational attainment, labor force status, and other characteristics, as well as by weekday and weekend day.

Survey Universe: ATUS covers all residents living in households in the United States that are at least 15 years of age, with the exception of active military personnel and people residing in institutions such as nursing homes and prisons.

Research Design: ATUS has a stratified three-stage sample. First, a sample is drawn from households that have completed their eighth and final month of interviews for the Current Population Survey (CPS), the monthly labor force survey in the United States. This sample is distributed across the States approximately equal to the proportion of the national population each one represents. Second, households are stratified based on the race or ethnicity of the householder, the presence and age of children, and the number of adults in adults-only households. Finally, an eligible person is randomly selected from each household; this individual becomes the designated person for an ATUS interview. Respondents are interviewed one time about how they spent their time on the previous day, where they were, and whom they were with.

Survey Mode: All ATUS interviews are conducted using Computer-assisted Telephone Interviewing. Procedures are in place to collect information from the small number of households that did not provide a telephone number during the final CPS interview.

Unit of Analysis: Person-level data analysis.

Sample: National probability sample designed to represent the U.S. civilian noninstitutionalized population ages 15 and over.

Topics: Time use.

Data Availability: ATUS is a continuous survey, with interviews conducted nearly every day of the year and a sample that builds over time. January 2003 marked the official beginning of ATUS data collection. Micro data files for 2003 and 2004, as well as supporting documentation, are available from the ATUS home page: <http://www.bls.gov/tus/>.

Linked Data: Current Population Survey.

Data Disseminated: ATUS 2003 to 2007 micro data files and supporting documentation are available for download from the ATUS home page: <http://www.bls.gov/tus/>.

Associated Reports: A list of survey documents and reports also is available on the ATUS home page: <http://www.bls.gov/tus/>.

Future Plans: Future ATUS micro data files will be released on an annual basis. In addition to annual releases of the core ATUS interview data, BLS plans to release files containing data collected in modules (a series of questions designed to obtain more detailed information about a time-use related topic) that are periodically added to the core survey for a limited length of time. The Economic Research Service (ERS) of the U.S. Department of Agriculture currently is sponsoring a Food & Eating module. The ERS module collects data on eating and drinking as a secondary activity, grocery shopping and food preparation patterns, food stamp eligibility, school meal programs, and height and weight. Release dates for this and future module files are yet to be determined.

For more information:

Email: [atusinfo@bls.gov](mailto:atusinfo@bls.gov)

Phone: 202-691-6339

Website: <http://www.bls.gov/tus/>

## **Consumer Expenditure Survey (CEX)**

Sponsor: Bureau of Labor Statistics

Data Collection Agency/Organization: U.S. Census Bureau

Purpose: To provide a continuous and comprehensive flow of data on the buying habits of American consumers for use in a wide variety of economic research activities and in support of periodic revisions of the Consumer Price Index.

Description: The CEX consists of two separate components: (1) a quarterly Interview panel survey in which expenditures of consumer units are obtained in five interviews conducted every 3 months, and (2) a diary or recordkeeping survey completed by participating consumer units for two consecutive 1-week periods. Each component has its own questionnaire and independent sample. The Interview sample is selected on a rotating panel basis with approximately 5,000 consumer units participating each quarter. For the Diary survey, approximately 5,000 consumer units participate each year. The CEX asks the age of respondents, and expenditures are tabulated by age and income.

Survey Universe: The surveys target the total noninstitutionalized population (urban and rural) of the United States in 1980, 1984, and thereafter. The surveys targeted the urban noninstitutionalized population in 1981 through 1983.

Research Design: In the Interview Survey, each consumer unit is interviewed every 3 months over five calendar quarters. In the initial interview, information is collected on demographic and family characteristics and on the consumer unit's inventory of major durable goods. Expenditure information also is collected in this interview, but is used only to prevent duplicate reporting in subsequent interviews. Expenditure information is collected in the second through the fifth interviews using uniform questionnaires. Income and employment information is collected in the second and fifth interviews. In the fifth interview, a supplemental section is administered in order to account for changes in assets and liabilities over a 1-year period. In the Diary Survey, respondents are asked to keep track of all their purchases made each day for two consecutive 1-week periods. Participants receive each weekly diary during a separate visit by a Census Bureau interviewer.

Survey Mode: Personal interviews and self enumerated forms.

Unit of Analysis: Person-level analysis.

Sample: National probability sample of households designed to represent the total noninstitutional civilian population.

Topics/Key words: Buying habits.

Data Availability: Annual income and expenditures integrated from the Interview and Diary surveys in varying detail, classified by income, age, consumer unit size, and other demographic

characteristics of consumer units, since 1984. Annual income and expenditures from the Interview and Diary surveys by selected consumer unit characteristics since 1980.

Linked Data: N/A

Data Dissemination: Data in several different formats are available from the Consumer Expenditure Survey website <http://www.bls.gov/cex/>. The standard data releases include the following:

- NEWS RELEASE. This annual release consists of a brief discussion of the latest survey results. The information generally is available on the Consumer Expenditure Survey website on the day on which the annual data are released. The annual data usually are released late in the year following the reference year (for example, 2002 data will be available late in 2003).
- ANNUAL REPORT. The report includes integrated data from the Diary and Interview portions of the Consumer Expenditure Survey in 10 standard tables. The tables show average expenditures, income, and characteristics for consumer units classified by 13 standard characteristics—quintiles of income, before-tax income class, age, size of the consumer unit, composition of the consumer unit, number of earners, housing tenure, race, type of area (urban or rural), Hispanic origin of reference person, region, occupation, and education.
- TWO-YEAR REPORTS. Two following separate reports are published in alternating years: (1) Biennial Report—this report includes integrated survey data and is published at 2-year intervals, (2) Anthology—this report includes both methodological and analytical articles.
- MICRODATA ON CD-ROMS. Diary and Interview Survey microdata—that is, data for individual consumer units—are available for purchase on CD-ROM. The interview files contain demographic and expenditure data.

Data/Reports: CEX data are available in news releases, reports, bulletins, articles in the Monthly Labor Review, and on tapes, diskettes, CD ROM, and the Internet.

For more information:

Email: [CEXINFO@bls.gov](mailto:CEXINFO@bls.gov)

Phone: 202-691-6900

Website: [www.bls.gov/cex/](http://www.bls.gov/cex/)

## **Current Population Survey (CPS)**

Sponsor: Bureau of Labor Statistics

Data Collection Agency/Organization: U.S. Census Bureau

**Purpose:** The Current Population Survey (CPS) provides monthly estimates of total employment, unemployment, and other characteristics of the civilian noninstitutional population 16 years old and over as well as for various demographic groups. The Annual Social and Economic Supplement (ASEC), formerly called the Annual Demographic Supplement (ADS), supplements the basic CPS labor force data with information on income, including noncash income sources such as food stamps, school lunch program, employer-provided group health insurance plan, employer-provided pension plan, personal health insurance, Medicaid, Medicare, CHAMPUS or military health care, and energy assistance. Data from the ASEC also includes information on the prior year's work experience of persons for whom information is collected including occupation, and industry.

**Survey Universe:** The survey universe is composed of persons 15 years of age and over in the civilian noninstitutional population. Published labor force data from the CPS are for those aged 16 years and over. While active-duty members of the Armed Forces are not asked questions regarding their labor force status, they are asked questions about their income.

**Research Design:** The basic CPS has been conducted since 1945, although some data were collected prior to that time. Collection of income data began in 1948. Over the years, the number of income questions in the ASEC has expanded. In 1994 major changes to the basic CPS labor force questions 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.

**Survey Mode:** Households in the sample are interviewed for 4 consecutive months, not interviewed for 8 consecutive months, and then interviewed again for 4 consecutive months (then dropped out of the sample). Over the whole 16-month period a household is interviewed eight times. The CPS includes uses both in-person and telephone interviews.

**Unit of Analysis:** Households, families, and persons.

**Sample:** The CPS sample is located in 754 sample areas, with coverage in every State and the District of Columbia. The basic CPS sample is selected from multiple frames using multiple stages of selection. Each unit is selected with a known probability to represent similar units in the universe. The sample design is a State-based design, with the sample in each State being independent of the others.

**Topics/Key Words:** Labor force, unemployment, demographic data.

Data Availability: CPS data can be obtained from either the Census Bureau (<http://www.census.gov/>) or the Bureau of Labor Statistics (<http://www.bls.gov/cps>).

Linked Data: See description of National Longitudinal Mortality Study.

Reports: U.S. Census Bureau. Technical Paper 63RV. Current Population Survey: Design and Methodology. TP63RV, March 2002 found at <http://www.census.gov/prod/2002pubs/tp63rv.pdf>.

Fore more information:

Website: [www.census.gov/cps/](http://www.census.gov/cps/)

## National Longitudinal Surveys (NLS)

Sponsor: Bureau of Labor Statistics

Data Collection Agency/Organization: Ohio State University Center for Human Resource Research and the National Opinion Research Center at the University of Chicago. Data for the original cohorts were collected by the U.S. Census Bureau.

Purpose: The National Longitudinal Surveys (NLS) are a set of surveys sponsored by the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor. These surveys have gathered information on the labor market activities and other significant life events of several groups of men and women. The same respondents have been interviewed repeatedly to provide a picture of schooling, work history, retirement preparation, and other events and transitions over the life cycle. For nearly 4 decades, NLS data have served as an important tool for economists, sociologists, demographers, and other researchers.

**Table 1. Summary of the National Longitudinal Survey Cohorts**

<b>Cohort</b>	<b>Birth Years</b>	<b>Year of first interview (and age at first interview)</b>	<b>Number of interview rounds completed</b>	<b>Original sample size</b>	<b>Status</b>
<b>Older Men</b>	4/1/1906 to 3/31/1921	1966 (45-59)	13	5,020	Ended in 1990
<b>Mature Women</b>	4/1/1922 to 3/31/1937	1967 (30-44)	21	5,083	Ended in 2003
<b>Young Men</b>	4/1/1941 to 3/31/1952	1966 (14-24)	12	5,225	Ended in 1981
<b>Young Women</b>	1943-1953	1968 (14-24)	22	5,159	Ended in 2003
<b>NLSY79</b>	1957-1964	1979 (14-22)	21	12,686 (9,964 remain eligible)	Continuing biennially
<b>NLSY79 Child</b>	Birth to age 14	1986	10	Varies	Continuing biennially
<b>NLSY79 Young Adult</b>	15 and over	1994	6	Varies	Continuing biennially
<b>NLSY97</b>	1980-1984	1997-98 (12-17)	8 (round 9 ends summer 2006)	8,984	Continuing biennially

The four original NLS cohorts began in the mid-1960s. They are called the NLS of Older Men, Mature Women, Young Men, and Young Women. Screening for these surveys was conducted in early 1966 by the Census Bureau. All residents of a screened household who were eligible for one of the cohorts were included, so the cohorts together include many spouse, parent-child, and



sibling combinations. The cohorts of Older Men, Mature Women, and Young Women are particularly useful for conducting research on aging.

#### Older Men

The NLS of Older Men included men who were ages 45–59 when the survey began in 1966. These men were well into their careers and on the threshold of decisions about the timing and extent of their labor force withdrawal. Data collection focused on work experiences, including job changes, retirement expectations and experiences, and reentry to the labor market after initial retirement.

The survey also included questions about retirement planning, health, insurance coverage, and the ways in which respondents spent their leisure time. Regular interviews with this cohort ceased in 1981. In 1990, respondents were interviewed a final time along with widows or other family members of the deceased sample members.

#### Mature Women and Young Women

The NLS of Mature Women included women who were ages 30–44 when the survey began in 1967. Many of these women were re-entering the workforce following childbearing and were balancing the roles of mother and labor force participant. Respondents were ages 66–81 when interviewed for a final time in 2003.

The NLS of Young Women included women who were ages 14–24 when the survey began in 1968. These women were completing school, making initial career decisions, and starting families. Respondents were ages 49–60 when interviewed for a final time in 2003.

Beginning in 1993, surveys of the mature and young women's cohorts included questions about transfers of time and money to a respondent's parents or children. These questions were added to obtain information about the financial impact aging parents have on their children and to record inheritances and other transfers that respondents received from their parents.

The NLS of Mature Women obtained information in all but two rounds from 1977 to 2003 on respondents' retirement plans, expectations, and eligibility for various pension plans. The NLS of Young Women included questions in 1991 and 1995–2003 on pension coverage and vesting rights for respondents who indicated that their employer provided a pension or other retirement plan. Both surveys of women also included a comprehensive set of health-related questions.

#### NLSY79

The National Longitudinal Survey of Youth 1979 (NLSY79) includes men and women who were born in the years 1957–64 and living in the United States when the survey began in 1979. The sample members, who constitute part of the baby boom generation, were ages 14–22 during the first round of data collection. They were ages 41–49 during the 22nd round of data collection in 2006. The NLSY79 was conducted annually from 1979 through 1994 and has been conducted biennially since 1994. A primary focus of the NLSY79 is labor force behavior, but the content of the survey is considerably broader. The NLSY79 includes questions on educational attainment, training, income and assets, participation in government programs, health, workplace injuries, insurance coverage, alcohol and drug use, sexual activity, marital and fertility histories,

and other topics. As the sample members soon enter their 50s, the NLSY79 will become increasingly valuable as a research tool on aging.

The NLSY79 Child and Young Adult surveys obtain a wealth of information on the children born to female NLSY79 respondents. The collection of data on these children began in 1986 with a battery of cognitive, social, emotional, and physiological assessments administered to NLSY79 children and their mothers. These biennial assessments are administered primarily in person. Beginning in 1988, children age 10 and over have answered a self-administered set of questions about family, friends, jobs, school, after-school activities, religious attendance, smoking, alcohol and drug use, and more. Starting in 1994, children who reach age 15 by December 31 of the survey year complete a questionnaire that is similar to the main NLSY79 and asks about work experiences, training, schooling, health, fertility, attitudes, and work expectations. This “Young Adult” interview, which is conducted primarily by telephone, replaced the child assessments for older adolescents. The data collected about the children and young adults can be linked easily with information collected from their mothers in the main NLSY79.

#### NLSY97

The National Longitudinal Survey of Youth 1997 (NLSY97) includes men and women born during the years 1980 through 1984. Survey respondents were ages 12–17 when first interviewed in 1997 and were ages 20–26 during the 9th round of interviews in 2005–06. The survey covers a variety of topics, including labor market status and characteristics of jobs, education, training, aptitudes, health, fertility, marital history, income and assets, participation in government programs, attitudes, sexual activity, criminal and delinquent behavior, household environment, and military experiences. Sample members today are obviously much too young for the NLSY97 to contribute to current research on aging. If the survey is able to continue successfully, however, decades from now it will be a valuable tool for aging-related research.

#### Data Availability:

The easiest way to obtain NLS data is to visit the NLS website and click on the “Obtaining NLS Data” link. From there, you can download public-use data files for free or order a multicohort CD for a small fee. The website also provides information on how to obtain access to geographic variables and other restricted-access NLS data.

For more information:

Email: [NLS\\_INFO@bls.gov](mailto:NLS_INFO@bls.gov)

Phone: 202-691-7410

Website: <http://www.bls.gov/nls/home.htm>

## **Department of Veterans Affairs**

### **National Survey of Veterans (NSV)**

Sponsor: Office of Policy, Assistant Secretary for Policy and Planning, Department of Veterans Affairs.

Data Collection Agency/Organization: Department of Veterans Affairs.

Purpose: The 2009 National Survey of Veterans (NSV) will satisfy the requirements detailed in P.L. 108-454, section 805. Information will be collected on awareness, demographics, health care, disability, life insurance, burial benefits, employment, education and training, and vocational rehabilitation. It will survey active duty service members, activated National Guard and Reserve members, and survivors and family members in addition to veterans.

Current Activities: Survey design in progress.

Future Plans: Execution and completion by the end of December 2009.

Public-Use Data Files: Survey results on a public-use data tape – fall 2010

For more information:

Contact: Maribel Aponte

Email: [Maribel.Aponte@va.gov](mailto:Maribel.Aponte@va.gov)

Phone: 202-461-5790

Website: <http://www1.va.gov/vetdata/> (click on “Surveys and Studies”)

Note that this website contains results from the 2001 National Survey of Veterans.

When results from the 2009 National Survey of Veterans become available, it will be posted on the website.

## **Survey of Veteran Enrollees' Health and Reliance Upon VA, 2007**

Sponsor: Office of the Assistant Deputy Under Secretary for Health (ADUSH) for Policy and Planning, Department of Veterans Affairs

Data Collection Agency/Organization: ORC Macro, Inc.

Purpose: The 2007 Survey of Veteran Enrollees' Health and Reliance Upon VA is the sixth in a series of surveys of veteran enrollees for VA health care conducted by the Veterans Health Administration (VHA), within the Department of Veterans Affairs (VA), under multiyear OMB authority. Previous surveys of VHA-enrolled veterans were conducted in 1999, 2000, 2002, 2003, and 2005. All six VHA surveys of enrollees consisted of telephone interviews with stratified random samples of enrolled veterans. In 2000, 2002, 2003, 2005 and 2007 the survey instrument was modified to reflect VA management's need for specific data and information on enrolled veterans.

As with the other surveys in the series, the 2007 Survey of Veteran Enrollees' Health and Reliance Upon VA sample was stratified by Veterans Integrated Service Network, enrollment priority, and type of enrollee (new or past user). Telephone interviews averaged 15 minutes in length. In 2007, the survey was conducted over the summer, with 42,587 enrollees participating.

VHA enrollee surveys provide a fundamental source of data and information on enrollees that cannot be obtained in any other way except through surveys and yet are basic to many VHA activities. The primary purpose of the VHA enrollee surveys is to provide critical inputs into VHA Health Care Services Demand Model enrollment, patient, and expenditure projections, and the Secretary's enrollment level decision processes; however, data from the enrollee surveys find their way into a variety of strategic analysis areas related to budget, policy, or legislation.

VHA enrollee surveys provide particular value in terms of their ability to help identify not only who VA serves but also to help supplement VA's knowledge of veteran enrollees' sociodemographic, economic, and health characteristics, including household income, health insurance coverage status, functional status (ADL and IADL limitations), perceived health status, race and ethnicity, employment status, smoking status, period of service and combat status, other eligibilities and resources, their use of VA and non-VA health care services and "reliance" upon VA, and their potential future use of VA health care services. The next planned VHA enrollee survey is for late 2009.

For more information:

Contact: Mary Beth Matthews

Email: [marybeth.matthews@va.gov](mailto:marybeth.matthews@va.gov)

Phone: 414-384-2000, ext. 42359

Website: <http://www1.va.gov/vhareorg/>

## **Veteran Population Estimates and Projections (model name is VetPop2007, Dec. 2007)**

Sponsor: Office of Policy, Assistant Secretary for Policy and Planning, Department of Veterans Affairs.

Data Collection Agency/Organization: Department of Veterans Affairs.

Purpose: To provide estimates and projections of the veteran population by age groups and other demographic characteristics at the National, State and County levels.

Description: Veteran estimates and projections were computed using a cohort-component approach, whereby 2000 Census baseline data were adjusted forward in time on the basis of separations from the armed forces (new veterans) and expected mortality.

Current Activities: Research methodologies to improve Migration and Mortality.

Future Plans: Incorporate data from the new VA DoD Identity Repository, and continue to research ways of using American Community Survey data.

Public-Use Data: National, State and county estimates and projections of veteran population by age groups and other demographic characteristics are available on VA's website.

Reports: See <http://www1.va.gov/vetdata/page.cfm?pg=15> for selected documents and special reports that use VetPop2007 data.

For More Information:

Contact: Cathy Tomczak

Email: [cathy.tomczak@va.gov](mailto:cathy.tomczak@va.gov)

Phone: 202-461-5769

Website: <http://www1.va.gov/vetdata/> (click on "Demographics")

## **Federal Reserve System, Board of Governors**

### **Survey of Consumer Finances (SCF)**

**Sponsor:** The study is sponsored by the Federal Reserve Board in cooperation with the Department of the Treasury.

**Data Collection Agency/Organization:** Since 1992, data have been collected by the National Organization for Research at the University of Chicago (NORC).

**Purpose:** The Survey of Consumer Finances (SCF) is a triennial survey of the balance sheet, pension, income, and other demographic characteristics of U.S. families. The survey also gathers information on the use of financial institutions. The results of the survey are widely used by researchers and policymakers to examine the effects of economic policies and trends on American households.

**Survey Universe:** Civilian noninstitutionalized population.

**Research Design:** The SCF is a triennial, nationally representative survey using a dual frame sample design. One part of the sample is a standard multistage national area probability sample. The second part of the sample employs information from SOI, under stringent provisions to protect the privacy of taxpayers, to select a sample with disproportionate representation of families more likely to be relatively wealthy; this sample is stratified by a “wealth index” computed using observed capital income flows and related information. The two parts of the sample are adjusted for sample nonresponse and combined using weights to provide a representation of families overall.

**Survey Mode:** CAPI

**Unit of Analysis:** Family

**Topics/Key Words:** Income, Wealth, Pensions

**Linked Data:** N/A

**Data Dissemination:** Public versions of the data from the 1962-63, 1983, 1986, 1989, 1992, 1995, 1998, 2001, 2004 cross-sections and the 1983-89 panels are available via the Internet at <http://www.federalreserve.gov/pubs/oss/oss2/scfindex.html>

**Reports:** Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances. Brian K. Bucks, Arthur B. Kennickell, and Kevin B. Moore, Federal Reserve Bulletin, vol. 92 (February 2006), pp. A1-A38.

**Working papers:**

Currents and Undercurrents: Changes in the Distribution of Wealth, 1989-2004  
Arthur B. Kennickell, Federal Reserve Board

Disentangling the Importance of the Precautionary Saving Motive  
Arthur B. Kennickell, Federal Reserve Board, Annamaria Lusardi, Dartmouth College and  
NBER, November 2005

Managing Data Quality on the 2004 Survey of Consumer Finances  
Leslie A. Athey, NORC, Arthur B. Kennickell, Federal Reserve Board, June 2005

Consider the Source: Differences in Estimates of Income and Wealth from Survey and Tax Data  
Barry Johnson, Internal Revenue Service, Kevin Moore, Federal Reserve Board, January 2005

The Good Shepherd: Sample Design and Control for Wealth Measurement in the Survey of  
Consumer Finances, Arthur B. Kennickell, Federal Reserve Board

Darkness Made Visible: Field Management and Nonresponse in the 2004 SCF  
Arthur B. Kennickell, Federal Reserve Board, August 2005

Future Plans: It is expected that the survey will be conducted every 3 years in the future.

For more information:

Contact: Gerhard Fries

Email: [m1gxf00@frb.gov](mailto:m1gxf00@frb.gov)

Phone: 202-452-2578

Contact: Arthur Kennickell

Email: [m1abk00@frb.gov](mailto:m1abk00@frb.gov)

Phone: 202-452-2247

Website: [www.federalreserve.gov/feedback.cfm](http://www.federalreserve.gov/feedback.cfm)

## Office of Management and Budget

### FedStats

Sponsor: Interagency Council on Statistical Policy

Purpose: FedStats (<http://www.fedstats.gov/>) is an interagency web portal that provides access to a full range of official statistical information available to the public from the Federal Government. It allows users to access official statistics collected and published by more than 100 Federal agencies without having to know in advance which agency produces them. FedStats users can utilize the Internet's powerful linking and searching capabilities to track economic and population trends, education, health care costs, aviation safety, foreign trade, energy use, farm production, and more.

FedStats, developed under the auspices of the Interagency Council on Statistical Policy (ICSP), permits easy access via an initial point of entry to the wide array of Federal statistics available to the public. FedStats provides a centralized set of links to the Internet sites that individual agencies maintain for disseminating Federal statistics. The site's primary objective is to help users find the information they need without having to know and understand in advance how the decentralized Federal statistical system is organized or which agency or agencies may produce the data they are seeking. FedStats has logged nearly 22 million user sessions since its release to the public in 1997.

The ICSP's Task Force on One-Stop Shopping for Federal Statistics continues to upgrade and expand FedStats' coverage of the government's statistical products. A "MapStats for Kids" section was developed and made available to the public to facilitate the improvement of statistical literacy for young children. The addition to the MapStats section of the site of statistical information for cities with populations of 25,000 or more has been well received by the public, and provides links to the Department of Housing and Urban Development's State of the Cities website for more in-depth statistical information for a locality. FedStats plans to refresh the basic MapStats data during the coming year. Search capabilities continue to be enhanced by indexing the statistics available on nearly all the FedStats agency websites, and exploratory work in the use of the Statistical Data and Metadata Exchange (SDMX) protocol to facilitate updates to information on the site is being conducted. The ICSP task force continues to respond to user requests for a broader scope of subjects, more detailed data on those subjects, and easier overall access to the data, and plans to further improve the public's access to statistical information through revision of the current set of "Topics A-Z" displayed on the site.

Note: Essentially all of the statistical information available through FedStats is maintained and updated solely by Federal agencies on their own web servers.

FedStats provides information for Federal agencies reporting expenditures of at least \$500,000 per year in one or more statistical activities including:



- Planning of statistical surveys and studies, including project design, sample design and selection, and design of questionnaires, forms, or other techniques of observation and data collection.
- Training of statisticians, interviewers, or processing personnel.
- Collection, processing, or tabulation of statistical data for publication, dissemination, research, analysis, or program management and evaluation.
- Publication or dissemination of statistical data and studies.
- Methodological testing or statistical research.
- Data analysis.
- Forecasts or projections that are published or otherwise made available for government-wide or public use.
- Statistical tabulation, dissemination, or publication of data collected by others.
- Construction of secondary data series or development of models that are an integral part of generating statistical series or forecasts.
- Management or coordination of statistical operations.
- Statistical consulting or training.

For More Information:

Website: [www.fedstats.gov](http://www.fedstats.gov)

## **Social Security Administration**

### **New Beneficiary Data System (NBDS)**

Sponsors: Social Security Administration, National Institute on Aging, Health Care Financing Administration, Assistant Secretary for Planning and Evaluation in HHS, Office of the Assistant Secretary for Health, Agency for Health Care Policy and Research

Data Collection Agency/Organization: Social Security Administration

Purpose: The New Beneficiary Data System (NBDS) is an increasingly important source of information on the changing circumstances of aged and disabled beneficiaries. Based initially on a survey of new beneficiaries conducted in 1982, the dataset was subsequently enhanced with information from administrative records, and in 1991, through follow-up interviews with survivors from the original survey. The result is a longitudinal data file—a rich source of information on these cohorts of aged and disabled persons during the first decade after they began receiving benefits.

#### **Description:**

##### **The New Beneficiary Survey (NBS)**

The New Beneficiary Survey was conducted in late 1982 with a sample representing nearly 2 million persons who had begun receiving Social Security benefits during a 12-month period in 1980–81. Interviews were completed with three types of beneficiaries: 9,103 retired workers, 5,172 disabled workers, and 2,417 wife or widow beneficiaries. (These sample numbers are slightly lower than the numbers previously reported in some publications. During the course of analysis, several hundred respondents were found to have received Social Security benefits prior to 1980–81. These respondents are excluded here and in later stages of the NBDS). In addition, interviews were obtained from 1,444 persons aged 65 or over who were covered by Medicare but were not receiving Social Security payments because of high earnings. The NBS interviews covered a wide range of topics, including demographic characteristics, marital and childbearing history, employment history, current income and assets, and health. Selected data were also gathered from spouses and added from administrative records.

##### **The New Beneficiary Follow-Up (NBF)**

Designed to meet a growing need for longitudinal data on the dynamics of disability and aging, follow-up interviews were conducted throughout 1991 with surviving respondents from the NBS. By the time of the 1991 survey, four-fifths of the original respondents were still alive. Interviews were completed with 87 percent of these survivors—3,428 of the original retired-worker men, 2,811 retired-worker women, and 1,754 wife or widow beneficiaries. By 1991, almost all of these persons were in their early- to mid-seventies.

#### **Survival Rates**

Preliminary data reveal that 24 percent of the original respondents had died by the time of the NBF. The wife or widow beneficiaries had the highest survival rate (84 percent), followed by the Medicare-only and retired-worker groups (79 percent and 78 percent, respectively).

Although disabled-worker beneficiaries were the youngest group, they had experienced the highest mortality, with only 69 percent surviving.

#### Links with Administrative Data

Since the 1982 survey, selected information on the NBS respondents has been compiled periodically from Social Security, Supplemental Security Income (SSI), and Medicare records. These administrative data—which can be linked to the earlier survey data—make it possible to analyze changes in respondents’ covered earnings, cash benefits, and health expenditures. The SSA annual administrative earnings extend from 1951 through 2000. Medicare expenditures from 1984 to 1992. SSA benefits, SSI benefits, and mortality status refer most recently to Spring 2001.

#### Questionnaire Topics:

The NBF questionnaire was designed with an emphasis on measuring changes over time. It updated the profile of economic circumstances obtained in the NBS, and added or expanded sections on health, family contacts, and post-retirement employment. The interviews also explored major changes in life circumstances that might underlie changes in economic status (for example, death of a spouse, an episode of hospitalization, or a change in residence). In addition, disabled-worker beneficiaries were asked about their efforts to return to work experiences with rehabilitation services, and knowledge of SSA work incentive provisions.

#### Data Availability:

The 1982 NBS and the 1991 NBF data are publicly available through the National Archive of Computerized Data on Aging. The NBF data can be linked to the public use NBS and administrative files.

The New Beneficiary Data System, the 1982 New Beneficiary Survey, the 1991 New Beneficiary Follow-up, and Administrative record data are available for downloading from the Social Security Administration at <http://www.ssa.gov/policy/docs/microdata/nbds/>.

#### Bibliography:

First findings from the NBS, with emphasis on the recent follow-up survey, have been presented in a series of brief statistical reports published by the Social Security Administration. The initial set—Notes 1 through 5—was published in the fall 1993 Social Security Bulletin. The first report provided an overview of the NBS data files, describing the original sample, the 1991 survivors, and the kind of information available in the NBS. Notes 2 through 5 focused on persons in the disabled-worker cohort, describing their death rates, recovery rates, health and functional status, work attempts, and marital status. The next set of reports—Notes 6 through 11—was published in the spring 1994 Bulletin. This set focuses on the cohort of aged beneficiaries in the NBS who first received Social Security benefits on the basis of old age in mid-1980–81. These notes examine the surviving aged cohort in 1991 in terms of their health and functional status, marital status, employment experience and earnings, income sources, changes in pensions, and their feelings of “well-being”.

Reports: SOCIAL SECURITY BULLETIN. (1) Statistical Notes for the NBF Disabled Population. Fall 1993 issue; (2) Statistical Notes for the NBF Aged Population. Spring 1994

issue; (3) Several articles on the NBS in the issues from 1984 through 1992. Also republished in SSA's 1982 New Beneficiary Survey—A Compilation of Reports.

For more information:

Contact: Howard M. Iams, Social Security Administration

Email: [howard.m.iams@ssa.gov](mailto:howard.m.iams@ssa.gov)

Phone: 202-358-6217

Website: [www.ssa.gov/policy/docs/microdata/nbds/](http://www.ssa.gov/policy/docs/microdata/nbds/)