Summary of Selected Medical Industry Data Sets

April2007

This resource was prepared by the University of Minnesota and is currently being updated by the Medical Industry Leader Institute at the Carlson School of Management.

www.csom.umn.edu/mili t:

Government Data Sets

Area Resource File (ARF) *1*

Behavioral Risk Factor Surveillance System (BRFSS) *2*

Center for Disease Control and Prevention (CDC) - Pregnancy Risk Assessment Monitoring System (PRAMS) *3*

Current Population Survey (CPS) *5*

Health and Retirement Survey (HRS) *6*

Healthcare Cost and Utilization Project (HCUP) *8*

Longitudinal Studies of Aging (LSOA) *10*

Medical Expenditure Panel Survey (MEPS) *11*

National Ambulatory Medical Care Survey (NAMCS) *12*

National CAHPS Benchmarking Database (NCBD) *13*

National Health Interview Survey (NHIS) *15*

National Home and Hospice Care Survey (NHHCS) *16*

National Hospital Ambulatory Medical Care Survey (NHAMCS) *17*

National Hospital Discharge Survey (NHDS) *18*

National Nursing Home Survey (NNHS) *19*

National Survey of Ambulatory Surgery (NSAS) *20*

Safety Net Monitoring Initiative *21*

State and Local Area Integrated Telephone Survey (SLAITS) Child Well-Being and Welfare Module *22*

State and Local Area Integrated Telephone Survey (SLAITS)- Health Module *23*

Survey of Income and Program Participation (SIPP) *24*

Government - Medicare Data Sets

Accessing CMS Limited Data Set Files *25*

Accessing CMS Research Identifiable Data *26*

Accessing Limited Data Set Files *27*

Data Available from Center for Medicare and Medicaid Services (CMS) *28*

Enrollment Database (EDB) Files *33*

Medicare Current Beneficiary Survey (MCBS) *34*

Medicare Health Outcomes Survey (HOS) *35*

Medicare Physician Identification and Eligibility Registry (MPIER) *37*
Medicare Provider Analysis and Review Files (MedPAR) *38*
National Claims History (NCH) Nearline File *39*
National Practitioner Data Bank (NPDB) *40*
Public Use Files *41*
Request for CMS Research Identifiable Data *42*
Standard Analytical Files (SAF) *43*

Public Data Sets

American Hospital Association (AHA) Annual Survey *45*
American Medical Association (AMA) *46*
Atlantic Information Services, Inc. (AIS) *49*
Community Tracking Study (CTS) *50*
Dartmouth Atlas of Health Care *51*
Dorenfest *52*
Employer Health Benefits Survey *53*
International City / County Management Association's (ICMA) *54*
InterStudy *55*
Joint Commission on Accreditation of Healthcare Organizations (JCAHO) *57*
Kaiser Data Request Form *58*
Medical Group Management Association (MGMA) *59*
Medical Records Institute *61*
National Health Information (NHI) *62*
National Survey of Alcohol, Drug, and Mental Health Problems (NSADMHP) *63*
National Survey of America's Families (NSAF) *64*
Panel Survey of Income Dynamics (PSID) *65*
Press Ganey *66*
Quality Compass *67*
Sherlock Company *69*
Solucient *71*

Government Data Sets

Area Resource File (ARF)

Overview: ARF is a county-specific health resources information system designed to be used by planners, policymakers, researchers, and other professionals interested in the nation's health care delivery system and factors that may impact health status and health care in the U.S. The basic county-specific Area Resource File is the nucleus of the overall ARF System. It is a database containing more than 7,000 variables for each of the nation's counties. ARF contains information on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. The ARF System is a synthesis of data from over 50 different primary source files. Quality Resource Systems maintains the ARF under contract by the National Center for Health Workforce Analysis (NCHWA), and the Bureau of Health Professions within the Health Resources and Services Administration.

Access: Limited data access is provided online at: http://www.arfsys.com/
Full data may be obtained on CD-ROM, magnetic tape, and 3480 Cartridges.

For additional information contact:

Quality Resource Systems, Inc.
11350 Random Hills Road, Suite 100
Fairfax, Virginia 22030-6044
Tel: 703-352-7393
Fax: 703-352-9024
E-mail: qrs@qrs-inc.com

Cost: $500 for ASCII file on CD-ROM, magnetic tape, or 3480 cartridge
$800 for Microsoft Access CD-ROM

Updated: Annually since 1980

Population: County level information for all 3,142 counties in the U.S.

Survey Method: No primary data collected. See individual primary data instrument for survey methods

Linkage: No linkage available

Behavioral Risk Factor Surveillance System (BRFSS)

Overview: The objective of the BRFSS is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population. Factors assessed by the BRFSS include tobacco use, health care coverage, HIV/AIDS knowledge or prevention, physical activity, and fruit and vegetable consumption. The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project of the Centers for Disease Control and Prevention (CDC), and U.S. states and territories.

Access: BRFSS data are available for public use via the Internet at the BRFSS Web site:
http://www.cdc.gov/brfss/technical_infodata/surveydata.htm

For additional information:

Centers for Disease Control and Prevention
National Center for Chronic Disease Prevention and Health Promotion
Mail Stop K-47
4770 Buford Highway, NE
Atlanta, GA 30341-3717
Tel: 770-488-2455
Fax: 770-488-8150
E-mail: cdcinfo@cdc.gov

Cost: Free (Publicly Held)

Updated: Annually from 1984 to present

Population: Variable by year; 247,964 records in 2002

Survey Methods: Survey is a questionnaire conducted exclusively via a computer-assisted telephone interview (CATI). Sample is collected using a disproportionate stratified sample design with telephone numbers being divided into three strata, which are sampled separately. The three strata; high, medium, and low density areas are sampled to obtain a probability sample of all households with telephones.

Linkage: No linkage available

Center for Disease Control and Prevention (CDC) - Pregnancy Risk Assessment Monitoring System (PRAMS)

Overview: PRAMS, the Pregnancy Risk Assessment Monitoring System, is a surveillance project of the Centers for Disease Control and Prevention (CDC) and state health departments. PRAMS collects state-specific, population-based data on maternal attitudes and experiences prior to, during, and immediately following pregnancy. The goal of the PRAMS project is to improve the health of mothers and infants by reducing adverse outcomes such as low birth weight, infant mortality and morbidity, and maternal morbidity. PRAMS provides state-specific data for planning and assessing health programs and for describing maternal experiences that may contribute to maternal and infant health.

Access: Requests for PRAMS data from multiple states are reviewed on an individual basis by CDC and the participating PRAMS states using a standard proposal format. Research proposals vary between 4-8 pages in length. It's recommended that every research proposal contain the following information: a title; background; methods; results; discussion; resources needed; project time frame; and references. If the proposal is approved, the CDC will create a SAS analysis data set for the primary researcher on a CD-ROM or a diskette(s). A letter describing the contents of the CD-ROM or diskette will accompany the data. Sample SAS and SUDAAN programs are available upon request.

Mail or send electronically research proposals to:

Denise D'Angelo, MPH
Applied Sciences Branch MS-K22
Division of Reproductive Health
Centers for Disease Control and Prevention
4770 Buford Hwy, NE
Atlanta GA 30341-3724
Email: DDAngelo@cdc.gov

Requests for PRAMS data for a single state should be directed to that state's PRAMS coordinator (see map on PRAMS Web Site (http://www.cdc.gov/reproductivehealth/srv_prams.htm) and click on the state of interest for contact information).

PRAMS data are not yet available for Minnesota – for more information contact: Cynthia Turnure (Tel: 651-296-6351).

Or for more information about the request/proposal process, please send an inquiry to ccdinfo@cdc.gov.

Cost: No charge

Updated: Currently information from 1988-2001 for participating states only.

Population: PRAMS not only provides state-specific data but also allows comparisons among participating states because the same data collection methods are used

in all states. Thirty-one states and New York City currently participate in PRAMS. Four other states previously participated. The PRAMS sample of women who have had a recent live birth is drawn from the state's birth certificate file. Each participating state samples between 1,300 and 3,400 women per year. Women from some groups are sampled at a higher rate to ensure adequate data are available in smaller but higher risk populations. Selected women are first contacted by mail. If there is no response to repeated mailings, women are contacted and interviewed by telephone. Data collection procedures and instruments are standardized to allow comparisons between states.

Survey methods: Selected women are first contacted by mail. If there is no response to repeated mailings, women are contacted and interviewed by telephone. Data collection procedures and instruments are standardized to allow comparisons between states. The PRAMS questionnaire includes core questions that are asked by all the states and state-specific questions that are chosen or developed by individual states.

Linkage: Multiple state files can be linked

Current Population Survey (CPS)

Overview: The CPS is a monthly survey that provides information on the labor force characteristics of the U.S. population. Data obtained from the CPS include: employment, unemployment, earnings, hours of work, and a variety of demographic characteristics. Supplemental questions provide data on a variety of topics including school enrollment, income, previous work experience, health, employee benefits, and work schedules. Survey is broken into three basic parts: (1) household and demographic data, (2) labor force information, and (3) supplemental information in months that include supplementary information. CPS's March Supplement is the most commonly used data source for estimating rates of uninsurance. The CPS is jointly sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics.

Access: Data files are available at CPS Web site:
http://www.bls.census.gov/cps/datamain.htm
Access via FTP or DataFerrett

Additional contact information:
http://stats.bls.gov/cps/cps_cont.htm
E-mail: cpshelp@info.census.gov

Cost: CPS data are free and publicly available

Updated: Annually with monthly supplements since 1940

Population: 57,000 housing units interviewed monthly
112,000 persons 15 years old and over
Demographic data for 31,000 children 0-14 years old
Sample drawn from 792 sample areas with coverage in every state

Survey Methods: Multistage stratified sample
First stage of sampling divides the U.S. into primary sampling units (PSU)
PSU's are grouped into strata to provide uniformity
Second stage draws sample of housing units within PSU
Housing units are rotated and surveyed for 4 consecutive months
Survey conducted with help of computer-aided personal interviews (CAPI) and telephone interviews (CATI)

Linkage: Currently no linkage available

Health and Retirement Survey (HRS)

Overview: The HRS is a comprehensive national longitudinal survey conducted by the Institute for Social Research at the University of Michigan and sponsored by the National Institute on Aging. The HRS companion survey, the Study of Assets and Health Dynamics Among the Oldest Old (AHEAD) was merged with the HRS in 1998 with respondents from each forming a cohort in a combined interview. At that same time, two new cohorts were added: the Children of the Depression Era (CODA), born in 1924-30 and War Babies (WB), born in 1942-47. The HRS is intended to provide data for researchers, policy analysts, and program planners who are studying decisions that affect retirement, health insurance, saving and economic well-being. One important focus of the HRS concerns the implications of the aging American population in terms of health and economic well being during the latter part of life. The HRS also considers the economic well being of those supporting older family members and public programs such as Social Security, Medicare and Medicaid. Data from the HRS supplies longitudinal data for researchers to begin to describe the trends of the economic, health and family status of Americans over age 50.

Access: All publicly available data can be accessed from the HRS website: http://hrsonline.isr.umich.edu/

Much of the data available has restricted use due to its identifying nature, such as Social Security Administration data, Zip Code data, pension data, Medicare claims data, and National Death Index data. Access to this data requires a rigorous review process.
A flowchart describing the review process is available at: http://hrsonline.isr.umich.edu/rda/

For additional information contact:

Health and Retirement Study
Survey Research Center
Institute for Social Research
University of Michigan
426 Thompson Street
Ann Arbor, MI 48104
Tel: 734-936-0314
Fax: 734-647-1186

Cost: Public and restricted data are free.

Updated: New survey information is collected biannually, however new data are released annually. For a timeline and detailed description of data collection timelines see: http://hrsonline.isr.umich.edu/intro/dataflow.html

Population: Variable by year: HRS began with 12,600 persons in 7,600 households; AHEAD began with 7,447 persons

Survey Methods: The HRS core sample design is a multistage area probability sample

of households. The HRS design includes three oversamples, which are introduced as supplements to the core national sample and are designed to increase the numbers of Black and Hispanic HRS respondents as well as the number of HRS respondents who are residents of the state of Florida.

Linkage:

Medicare files
National Death Index
Social Security Administration data
Employer Pension Study

Healthcare Cost and Utilization Project (HCUP)

Overview: HCUP data collections are used for research on hospital utilization, access, charges, quality and outcomes at the national, regional, and state levels. The data are also usefully applied to patterns of care for diseases as well as for the study of population sub-groups such as minorities, children, women, and the uninsured. The survey is based on data collected by individual participating states and maintained by the Agency for Healthcare Research and Quality (AHRQ). Databases include: State Inpatient Databases (SID), Nationwide Inpatient Sample (NIS), the Kid's Inpatient Database (KID), State Ambulatory Surgery Databases (SASD), and in pilot phase the State Emergency Department Data (SEDD).

Access:

NIS & KID public data are available at HCUPnet:
www.ahrq.gov/data/hcup/hcupnet.htm
All data available on CD-ROM

For additional information:

HCUP Central Distributor, Social and Scientific Systems, Inc.
Tel: 866-556-4287
E-mail: hcup@s-3.com

Cost:

Data available on HCUPnet is free and publicly available
CD-ROMs are available for purchase
Costs vary and range from $160-$320 per database per year
User agreements are required for all non HCUPnet data

Updated:

SID	Annually from 1995
NIS	Annually from 1988
KID	1997 and 2000
SASD	Annually from 1997
SEDD	in pilot phase and currently unavailable

Population:

SID	33 states report representing 80% of all U.S. discharges, totaling 29 million inpatient discharge abstracts
NIS	drawn from SID approximates 20% of U.S. hospitals and includes 7 million inpatient discharges from 986 hospitals
KID	27 states reporting 1.9 million hospital inpatient discharges of adolescents from 2,500 hospitals
SASD	18 participating states

Survey Methods:

SID	contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. Contain 100 clinical and non-clinical variables.
NIS	a stratified probability sample of hospitals drawn from the SID
KID	adolescent specific subset of SID data

SASD contain the ambulatory surgery encounter abstracts in participating States, translated into a uniform format to facilitate multi-state comparisons and analyses. Contain 100 clinical and non-clinical variables.

Linkage: Certain NIS reporting hospitals data may be cross-linked with the Annual Survey of the American Hospital Association

Longitudinal Studies of Aging (LSOA)

Overview: The LSOA is a multi-cohort study of persons 70 years of age and over designed primarily to measure changes in the health, functional status, living arrangements, and health services utilization of two cohorts of Americans as they move into and through the oldest ages. The project is comprised of four surveys: the 1984 Supplement on Aging (SOA), the 1984-1990 Longitudinal Study of Aging (LSOA), the 1994 Second Supplement on Aging (SOA II), and the 1994-2000 Second Longitudinal Study of Aging (LSOAII). Baseline data are collected with the SOA and SOAII, supplements to the National Health Interview Survey (NHIS). The LSOA is comprised of three follow-up interviews and the LSOAII is comprised of four follow-up interviews. The study is a collaborative project of the National Center for Health Statistics (NCHS) and the National Institute on Aging (NIA). The study acquired initial data via supplemental questionnaires to the National Health Interview Survey. After identification of the sample cohort, the study tracks individuals for six years with two follow-up interviews.

Access: All data are available as ASCII or ASCII with SAS read-in capability files. LSOA, SOA, and SOA II data are available on CD-ROM for purchase. The two follow-up interviews that comprise the LSOAII data are available publicly through the NCHS Web-site:
http://www.cdc.gov/nchs/about/otheract/aging/lsoa2.htm

For additional information or to purchase a CD-ROM contact:

Julie Dawson Weeks, Ph.D.
Division of Epidemiology,
Office of Analysis, Epidemiology and Health Promotion
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Mailstop 6226
Hyattsville, Maryland 20782
Tel: 301-458-4562
E-mail: lsoa@cdc.gov

Cost: Free (publicly held)

Updated: 1984-1990 and 1994-2000

Population:
SOA: sample of 16,148 people age 55+
LSOA: follow-up interview of 7,527 people age 70+
SOAII: sample of 9,447 people age 70+
LSOAII: follow-up interview of 9,447 people age 70+

Survey Methods: SOA and SOAII are probability samples with a multistage, stratified, cluster design. Sampling strategy includes oversampling of black and Hispanic subpopulations and were conducted with personal interviews and computer assisted personal interviews.

Linkage: Data can be linked to Medicare records, National Death index records, and multiple cause of death records

Medical Expenditure Panel Survey (MEPS)

Overview: MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. Survey is sponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS). Four components: Household Component (HC), Nursing Home Component (NHC), Medical Provider Component (MPC), Insurance Component (IC)

Access: MEPS data releases are available for public use on diskettes, CD-ROMs, and the Internet at the MEPS Web site: http://www.meps.ahrq.gov

Additional information and research specific questions available:

William Carroll
CCFS Data Center Coordinator
Division of Survey Operations, Center for Cost and Financing Studies
Agency for Healthcare Research and Quality
2101 East Jefferson Street
Rockville, MD 20852
Tel: 800-358-9295
E-mail: CCFSDC@AHRQ.GOV

Cost: Free (Publicly Held); except for NHC and MPC which requires $150 fee and user agreement

Updated: Annually from 1996 to present

Population:

HC	sample consists of ≈ 11,000 households
NHC	sample consists of ≈ 815 nursing homes and 6,000 residents
MPC	sample consists of ≈ 22,000 providers to supplement HC information
IC	annual survey of ≈ 40,000 employers, insurers, and unions

Survey Methods:

HC	overlapping survey panel design in which data are collected over two and a half years in a series of five rounds of interviewing. The Household survey collects data through computer-assisted personal interviewing (CAPI)
NHC	12 CAPI questionnaires and for community information via computer assisted telephone interviewing (CATI).
MPC	paper and pencil instrumentation using CATI and a self-administered questionnaire.
IC	includes six establishment level questionnaires, four supplemental sheets on which plan level information is collected, and three person-level questionnaires all utilizing CATI

Linkage: MEPS Household Component is a subset of the National Health Interview Survey (NHIS)

National Ambulatory Medical Care Survey (NAMCS)

Overview: NAMCS is a national survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States. Data obtained include; demographic characteristics, symptoms, diagnoses, services, medications, diagnostics, and future treatment. Findings are based on a sample of visits to office-based physicians who are primarily engaged in direct patient care. NAMCS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NAMCS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm#Micro-data
Additional data for research purposes is available after submission of a research proposal.

For more information:

Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: Annually 1989 to present

Population: 21,000-36,000 patient records from 2,500-3400 participating physicians

Survey Methods: Survey is conducted by physicians and overseen by the U.S. Bureau of the Census. NAMCS utilizes a multistage probability design survey that involves probability samples of primary sampling units (PSU), physician practices within PSU's, and patient visits within practices.

Linkage: NAMCS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

National CAHPS Benchmarking Database (NCBD)

Overview: The NCBD aggregates respondent-level data files across sponsors and health plans for the Commercial, Medicaid, and SCHIP populations. The NCBD includes two major data sets: CAHPS Survey Response Data and Survey Administration and Health Plan/Provider Characteristics Data.

1. **CAHPS Survey Response Data:** The core of the NCBD is composed of data from the CAHPS Health Plan Survey. These data are submitted at the individual respondent level by participating sponsors in accordance with specifications provided by the NCBD. These data include all information from the survey, except for respondent names, addresses, telephone numbers, and member ID numbers. The NCBD does not include any data that could be used to identify individual respondents. Certain survey administration data (e.g., mode of administration, survey language) and descriptive information (e.g., state, zip code) may also be included in this data set.

2. **Survey Administration and Health Plan/Provider Characteristics Data**
   This data set includes:
   - Information regarding survey administration, such as mode of administration, response rates, and dates of survey completion.
   - Descriptive information relating to each of the sampled units (e.g., health plan products or provider groups), such as type of organization, size of enrollment, tax status and ownership, and location.

   More information is available on the NCBD website at: http://ncbd.cahps.org/Home/index.asp

   For more information:
   Dale Shaller, NCBD Managing Director
   Email: d.shaller@comcast.net
   Ph: 651-430-0759

Access: To obtain authorized access to the NCBD research files, researchers must submit an application that is reviewed by a special committee of the NCBD Advisory Group. Researchers are invited to submit an application for specific analysis projects to the NCBD Executive Research Committee for review and approval. If you would like access to NCBD data files for research purposes, you must:

- Read the Data Release Policy
- Submit a written request that includes:
  - The title of the proposed research project
  - A statement of the problem and the objectives of the proposed research
  - Specific data files requested (e.g., commercial, Medicaid, adult, children, survey version, survey time period, etc.) and the specific substantive items that will be used in the analyses (e.g., respondent age, race/ethnicity, etc.)
  - A description of analytic methods
  - A description of how the data will be stored and protected
  - A description of products that will result from the research
  - Once you have received approval, submit a signed Data Use Agreement that commits you to specific procedures for safeguarding

the confidentiality of the data. The agreement also specifies a statement you must use to acknowledge the NCBD as a data source in published analyses.

The review period for data requests (i.e., the time from receipt of request to notification of ERC decision) is typically 2 to 3 weeks. To date, the NCBD has authorized access to the data files for all applicants.

Please send proposals for data access and signed data use agreements to:
Jenny Crabb.
Westat
1650 Research Boulevard, RA 1137
Rockville, MD 20850
Ph: 301-279-4504
Fax: 301-294-3928
Email: JennyCrabb@westat.com

Cost: NCBD is supported through contract funding from the Agency for Healthcare Research and Quality (AHRQ). There are no participation fees provided that data requests are submitted according to NCBD data specifications and submission deadlines. Participating sponsors receive a standard sponsor-specific Sponsor Report at no cost. Fees may be charged to cover the costs of custom analyses and technical assistance, and for compiling data files in response to authorized requests.

Updated: Annually

Population: The NCBD currently contains six years of data from the CAHPS Health Plan Survey. The 2003 database holds survey results for approximately 350,000 adults and children enrolled in Commercial, Medicaid, SCHIP, and Medicare plans. For 2000-2003 composition see:
http://ncbd.cahps.org/AboutNCBD/AboutNCBD.asp#purpose

Survey method: Mail and/or phone survey

National Health Interview Survey (NHIS)

Overview: The NHIS obtains information about the amount and distribution of illness, its effects in terms of disability and chronic impairments, and the kinds of health services people receive. Supplemental NHIS data provide information on topics such as AIDS, child health care and immunization, dental care, substance abuse, hospitalization, preventive care, nursing care, prosthetics, and self-care. Additionally, there are supplements on aging and disability. The NHIS is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: Data available on CD-ROM or downloaded via:
http://www.cdc.gov/nchs/nhis.htm and
http://www.cdc.gov/nchs/products/elec_prods/subject/nhis.htm - description
Additional data for research purposes is available after submission of a research proposal.

For more information:
Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held) for download; CD-ROM $300-$1,000 per year;
Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center

Updated: Annually since 1957

Population: Recent surveys consist of 39,000 households yielding data for 100,000 persons in all 50 states

Survey Methods: Face-to-face computer assisted personal interviews conducted by U.S. Bureau of Census employees
Survey is a probability sample with a multistage, stratified, cluster design
Sampling strategy includes oversampling of black and Hispanic Subpopulations

Linkage: Limited linkage to the Medical Expenditure Panel Survey (MEPS) and The National Death Index (NDI)

National Home and Hospice Care Survey (NHHCS)

Overview: NHHCS is a continuing series of probability surveys of home and hospice care agencies in the United States. Information was collected about agencies that provide home and hospice care and about their current patients and discharges. Data collected depict both the characteristics of these health care providers and the people they serve, for example; type of ownership and affiliation, Medicare and Medicaid certification, patient demographics and functional status, diagnoses, services received, types of service providers, patient living arrangements and caregiver; expected sources of payment; and reason for discharge. NHHCS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NHHS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/nhhcsd/nhhcsd.htm
Additional data for research purposes is available after submission of a research proposal.

For more information:

Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: Periodically from 1992 to 2000

Population: 1,800 home health or hospice agencies; 3,400-5,400 current residents; 3,000-4,900 discharges

Survey Methods: NHHCS utilized three questionnaires and two sampling lists to collect the data. The sample design for the NHHCS is a stratified two-stage probability design. The first stage consisted of the selection of a stratified sample of agencies. Data are obtained through personal interviews with agency administrators and staff primarily responsible for the sampled patients care. Respondents also refer to patient medical and other records, as necessary. The second stage of sample selection utilized a sample selection table to obtain systematic probability samples of current patients and discharges.

Linkage: NAMCS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

National Hospital Ambulatory Medical Care Survey (NHAMCS)

Overview: NHAMCS is designed to collect data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments. Data obtained include; demographic characteristics, expected source of payment, patient complaints, diagnoses, diagnostic/screening services, procedures, medication therapy, disposition, types of health care professionals seen, causes of injury where applicable, and certain characteristics of the hospital. Findings are based on a national sample of visits to hospital emergency and outpatient departments. NHAMCS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NHAMCS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm#Micro-data
Additional data for research purposes is available after submission of a research proposal.

For more information:
Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: Annually 1992 to present

Population: 600 hospitals representing 21,000-36,000 Emergency Department encounters and 29,000-35,000 outpatient visits

Survey Method: Survey is conducted by hospital staff and overseen by the U.S. Bureau of Census. NHAMCS utilizes a multi-stage probability design survey that involves primary sampling units (PSUs), hospitals within PSUs, clinics within hospitals, and patient visits within clinics.

Linkage: NHAMCS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

National Hospital Discharge Survey (NHDS)

Overview: NHDS is a national probability survey designed to meet the need for information on utilization and characteristics of inpatient discharges from short-stay hospitals. Data obtained include; patient demographics, medical diagnoses and procedures, expected sources of payment, length of stay, discharge information, and certain characteristics of the hospital. Findings are based on a national sample of visits to various size hospitals. NHDS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NHDS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm
Additional data for research purposes is available after submission of a research proposal.

For more information:

Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: Annually 1965 to present

Population: 270,000 inpatient records acquired from about 500 hospitals

Survey Methods: Since 1988 two data collection methods have been utilized. One is a manual system that utilizes hospital medical records to abstract necessary information. These data are collected and processed by either hospital staff or staff of the U.S. Bureau of the Census. The second data collection system is an automated system in which NCHS purchases machine-readable medical record data from commercial organizations, state data systems, hospitals, or hospital associations.

Linkage: NHDS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

National Nursing Home Survey (NNHS)

Overview: NHHS is a continuing series of national sample surveys of nursing homes that provides information from two perspectives, the provider of services and the recipient. Data about the facilities include; characteristics such as size, ownership, Medicare/Medicaid certification, occupancy rate, number of days of care provided, and expenses. For recipients, data are obtained on demographic characteristics, health status, and services received. NNHS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NHHS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/nnhsd/nnhsd.htm
Additional data for research purposes is available after submission of a research proposal.

For more information:
Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: Periodically from 1973 to 1999; survey has undergone a major redesign and will be fielded again in 2003

Population: 1,500 nursing home facilities; 5,200-8,200 current residents; 6,000-6,900 discharges

Survey Methods: NNHS utilized three questionnaires and two sampling lists: Facility Questionnaire, Current Resident Questionnaire, Discharged Resident Questionnaire, Current Resident Sampling List, and Discharged Resident Sampling List. The sampling was basically a stratified two-stage probability design. The first stage was the selection of facilities and the second stage was the selection of residents and discharges. The second-stage sampling of current residents and discharges was carried out by interviewers.

Linkage: NAMCS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

National Survey of Ambulatory Surgery (NSAS)

Overview: NSAS is a national survey designed to meet the need for information about the use of ambulatory surgery services in the United States. Data obtained include; patient demographics, patient disposition, total cost, surgery times, expected sources of payment, region of the country where procedure was performed, and diagnoses and procedures performed. Sampled facilities included those places where surgical and non-surgical procedures are performed on an outpatient basis, including hospital or freestanding center general operating rooms, dedicated ambulatory surgery rooms, and other specialized rooms such as, endoscopy units and cardiac catheterization labs. NSAS is a component of the National Health Care Survey (NHCS) and is conducted by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC).

Access: NSAS data releases are available for public use on CD-ROM or via the NHIS web site:
http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm
Additional data for research purposes is available after submission of a research proposal.

For more information:

Kenneth W. Harris
Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, Maryland 20782
Voice Mail: (301) 458-4277
Fax: (301) 458-4039
E-mail: rdca@cdc.gov

Cost: Free (Publicly Held); Data requiring submission of a research proposal is subject to pricing via the NCHS Research Data Center.

Updated: 1994-1996 and survey resumption currently being planned

Population: 120,000 patient visits in about 500 facilities

Survey Method: NSAS uses a multistage probability design with independent samples of hospitals and freestanding ambulatory surgery centers selected at the first and second stages and visits to these facilities selected at the final stage. The remaining sample of facilities is selected using a three-stage stratified cluster design. The first stage consists of a selection of a sub sample of the primary sampling units (PSU). PSU's are counties, a group of counties, towns or townships. The second stage consists of a selection of facilities from the sample PSU's. At the third stage, a systematic random sample of ambulatory surgery visits is selected.

Linkage: NHDS data are linked to the National Death Index (NDI) and the Medicare Expenditure Panel Survey (MEPS)

Safety Net Monitoring Initiative

Overview: In response to a 2000 Institute of Medicine Report that described America's health care safety net as "intact but endangered," the Agency for Healthcare Research and Quality (AHRQ) and the Health Resources and Services Administration (HRSA) combined to lead a joint safety net monitoring initiative. The Safety Net Monitoring Initiative provides a synthesis of data as collected by numerous separate primary data investigations. Examples include the U.S. Census, the Current Population Survey (CPS), the American Hospital Association Annual Survey, the Healthcare Cost and Utilization Project, and other state and federal data sources. Examples of data categories include; demand for safety net services, financial support for safety net services, safety net structure and health system context, community context, and outcomes and safety net performance.

Access: Data are available in numerous formats for download from:
http://www.ahrq.gov/data/safetynet/document.htm
Technical support is available via E-mail: safenet@ahrq.gov

For more information:
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
Tel: 301-427-1364

Cost: Free (Publicly Available)

Updated: 2003 with update anticipated in 2004

Population: 118 measures are available for 30 states, including 355 counties and 172 cities in 90 metropolitan areas, as well as all 1,818 counties (both metropolitan and non-metropolitan) in those states

Survey Methods: No primary data collected. See individual primary data instrument for survey methods

Linkage: No linkage available

State and Local Area Integrated Telephone Survey (SLAITS)
Child Well-Being and Welfare Module

Overview: The State and Local Area Integrated Telephone Survey (SLAITS) is a mechanism for government agencies and nonprofit organizations to sponsor data collection in areas ranging from health insurance coverage and access to care to perceived health status and utilization of services. SLAITS uses the same random-digit-dial telephone design approach and sampling frame as the ongoing National Immunization Survey conducted by the Centers for Disease Control and Prevention (CDC.) It allows researchers to collect data using customized questionnaires while utilizing the National Immunization Survey sampling frame of nearly one million households. There are presently four existing SLAITS survey modules including the Child Well-Being and Welfare Module (CWBW), which was conducted in Texas and Minnesota. This module is intended to target and analyze the issue of uninsured children. The primary goal of this survey was to evaluate health insurance data and describe the well being of children enrolled in federal Medicaid and state funded health insurance programs.

Access: Survey data (ASCII format with SAS input files available) is available at the National Center for Health Statistics Website: http://www.cdc.gov/nchs/about/major/slaits/child_tx_mn.htm

For additional information contact:

National Center for Health Statistics
Division for Health Interview Statistics
Attention: SLAITS
3311 Toledo Road, Room 2113
Hyattsville, MD 20782
Fax: (301) 458-4035
E-mail: slaits@cdc.gov

Cost: Free (Publicly held)

Updated: 1998 – 1999

Population: 2,192 children under the age of 18

Survey Methods: Survey is a probability sample with a multistage, stratified, cluster design. SLAITS is a list-assisted random-digit-dialing telephone survey. Interviews were conducted by NIS screeners utilizing computer assisted telephone interviews (CATI). Oversampling of low-income households with children (if necessary) to ensure that at least half the final sample consists of children living in households with income below 200% of the Federal Poverty Level.

Linkage: Select files can be linked to the National Immunization Survey (NIS)

State and Local Area Integrated Telephone Survey (SLAITS)- Health Module

Overview: The State and Local Area Integrated Telephone Survey (SLAITS) is a mechanism for government agencies and nonprofit organizations to sponsor data collection in areas ranging from health insurance coverage and access to care to perceived health status and utilization of services. SLAITS uses the same random-digit-dial telephone design approach and sampling frame as the ongoing National Immunization Survey conducted by the Centers for Disease Control and Prevention (CDC.) It allows researchers to collect data using customized questionnaires while utilizing the National Immunization Survey sampling frame of nearly one million households. There are presently four existing SLAITS survey modules including the Health Module that was conducted in Iowa and Washington States in 1997. The Health Module provides health and welfare related population based data at the State and local levels to track and monitor the health and well being of children and adults. Questions cover topics such as household composition, health care access and utilization, health status, health insurance, and income.

Access: Survey data (ASCII format) is available at the National Center for Health Statistics Website:
http://www.cdc.gov/nchs/about/major/slaits/mod_io_wa.htm

For additional information contact:
National Center for Health Statistics
Division for Health Interview Statistics
Attention: SLAITS
3311 Toledo Road, Room 2113
Hyattsville, MD 20782
Fax: (301) 458-4035
E-mail: slaits@cdc.gov

Cost: Free (Publicly held)

Updated: 1997

Population: Iowa: 2,675 persons; 1,021 households
Washington: 2,866 persons; 1,068 households

Survey Methods: Survey is a probability sample with a multistage, stratified, cluster design. SLAITS is a list-assisted random-digit-dialing telephone survey. Interviews were conducted by NIS screeners utilizing computer assisted telephone interviews (CATI).

Linkage: Select files can be linked to the National Immunization Survey (NIS)

Survey of Income and Program Participation (SIPP)

Overview: The Survey of Income And Program Participation (SIPP) is a longitudinal study conducted by the U.S. Census Bureau to collect information about the labor force behavior, income, participation in public programs, basic demographic characteristics to measure the effectiveness of existing federal, state, and local programs. In addition, the data are used to estimate future costs and coverage for government programs, such as food stamps, as well as to provide improved statistics on the distribution of income in the country. SIPP provides data on medical expenses and utilization of health care. SIPP is a good data set for analyzing the dynamics of the uninsured over a long period of time, which may provide a starting point for studies of the episodic nature of uninsurance. SIPP also helps fill the gaps that the Current Population Survey (CPS) leaves by providing data that allow a better understanding of the distribution of income, wealth, and poverty in the society, and of the effects of federal and state programs on the well-being of families and individuals.

Access: SIPP micro-data files are available in ASCII format and may be accessed through the Inter University Consortium for Political and Social Research (ICPSR)Web site:
http://www.icpsr.umich.edu:8080/ICPSR-SERIES/00135.xml

SIPP micro-data files are additionally available direct from the U.S. Census Bureau on CD-ROM. For additional information go to the Census Bureau Web-site on SIPP:
http://www.census.gov/mp/www/Tempcat/SIPP.html
or call 301-763-5263.

Cost: SIPP data obtained from the ICPSR Web-site is free
SIPP data obtained on CD-ROM from the Census Bureau is $50 per year.

Updated: Data gathered continuously since 1984. For the 1984-1993 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. Panel also expected in 2004.

Population: Sample size ranges from approximately 14,000 to 36,700 interviewed households per year.

Survey Methods: SIPP is a is a longitudinal, multistage-stratified survey primarily of adults in households in the United States, interviewed at least nine times at four-month intervals and followed over the life of the panel. The duration of each panel ranges from 2 1/2 years to 4 years.
Interviews are conducted by personal visit and by decentralized telephone.

Linkage: No linkage available

Government – Medicare Data Sets

Accessing CMS Limited Data Set Files

Overview: Although Limited Data Set (LDS) Files have been stripped of specified direct identifiers, the Centers for Medicare and Medicaid Services (CMS) still considers these beneficiary level files to be identifiable. As a result the data remain subject to the various privacy laws governing their distribution. Restrictions governing access and dissemination of LDS files are not as severe as requests for Research Identifiable Files (RIF). After submitting a data request and receiving approval from CMS requestors will not be required to get approval for subsequent LDS file requests, provided there is no change in how the data are being used. Requests for LDS files are typically honored within 6 to 8 weeks. Complete instructions for how to access LDS files is available on the CMS website:
http://www.cms.hhs.gov/data/purchase/default.asp

As part of a request for LDS data the following documentation is required:

- **Project Description**- this original document supplied by the requesting researcher provides information relating to the proposed project. Elements that are required include: research purpose, project issues and methods, data management safeguards, key personnel, implementation and dissemination of findings, and a statement of proprietary information (if applicable).

- **Data Use Agreement**- this legal document presents the requirements, policies, and procedures that the data requester is required to follow.

- **Data Request Form**- this simple form specifies which CMS data set is being requested, the requested format, and payment information.

Completed requests being sent via U.S. Postal Service should be submitted to:

Centers for Medicare and Medicaid Services
Public Use Files
Accounting Division
P.O. Box 7520
Baltimore, Maryland 21244-1850

Express mail should be submitted to:

Centers for Medicare and Medicaid Services
OFM / Division of Accounting-PUF
7500 Security Blvd., C3-07-11
Baltimore, Maryland 21244-1850

Accessing CMS Research Identifiable Data

Overview: Due to the confidential nature of Research Identifiable Files (RIF) requests for these data files involves extensive documentation and are subject to in-depth scrutiny by the Centers for Medicare and Medicaid Services (CMS). CMS no longer releases 100% files for RIF data. Data requests that are approved by CMS will only entitle the requesting researcher to those files that are relevant to the research project. Any alterations to the proposed research that necessitate obtaining additional RIF files requires the resubmission of a full request packet to CMS. The typical time frame for approval of requests for RIF data are between 3 and 6 months. Additional detailed information and access to the various forms is available on the CMS website. http://www.cms.hhs.gov/data/requests/academicandprivate.asp

As part of a request for RIF data the following documentation is required:

- **Written Letter of Request**- this letter outlines the purpose for the data request and includes a brief description of what data are required.

- **Data Use Agreement**- this legal document presents the requirements, policies, and procedures that the data requester is required to follow.

- **Data Request Form**- this simple form indicates which CMS data set is being requested and in what media format.

- **Privacy Board Review Summary Sheet**- Submission of this document indicates that the proposed research has IRB approval. This document must be completed by the researcher's Internal Review Board (IRB) and must be signed by an IRB board member.

- **Research Protocol**- this original document supplied by the requesting researcher provides detailed information relating to the proposed research. Elements that are required include: executive summary, objectives, background, importance, research question and methods, evaluation and analysis plan, work plan, staff and researcher qualifications, and implementation potential.

- **Evidence of Funding**- this form provides proof to CMS that the requester has the funds necessary to complete the research proposal.

If the proposed research is being federally funded, the researcher will also need to submit a letter of support from their federal project officer.

Completed request packets should be submitted to:

Robyn Thomas, Director
Division of Quality Coordination and Data Distribution (DQCDD)
OIS/EDG/DQCDD N1-15-03
Centers for Medicare and Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Accessing Limited Data Set Files

Overview: Although Limited Data Set (LDS) Files have been stripped of specified direct identifiers, the Centers for Medicare and Medicaid Services (CMS) still considers these beneficiary level files to be identifiable. As a result the data remain subject to the various privacy laws governing their distribution. Restrictions governing access and dissemination of LDS files are not as severe as requests for Research Identifiable Files (RIF). After submitting a data request and receiving approval from CMS requestors will not be required to get approval for subsequent LDS file requests, provided there is no change in how the data are being used. Requests for LDS files are typically honored within 6 to 8 weeks. Complete instructions for how to access LDS files is available on the CMS website:
http://www.cms.hhs.gov/data/purchase/default.asp

As part of a request for LDS data the following documentation is required:

- **Project Description**- this original document supplied by the requesting researcher provides information relating to the proposed project. Elements that are required include: research purpose, project issues and methods, data management safeguards, key personnel, implementation and dissemination of findings, and a statement of proprietary information (if applicable).

- **Data Use Agreement**- this legal document presents the requirements, policies, and procedures that the data requester is required to follow.

- **Data Request Form**- this simple form specifies which CMS data set is being requested, the requested format, and payment information.

Completed requests being sent via U.S. Postal Service should be submitted to:

Centers for Medicare and Medicaid Services
Public Use Files
Accounting Division
P.O. Box 7520
Baltimore, Maryland 21244-1850

Express mail should be submitted to:

Centers for Medicare and Medicaid Services
OFM / Division of Accounting-PUF
7500 Security Blvd., C3-07-11
Baltimore, Maryland 21244-1850

Data Available from Center for Medicare and Medicaid Services (CMS)

For additional comprehensive information on data available from CMS access the CMS Data Users Reference Guide (DURG) at: http://www.cms.hhs.gov/data/durg/ or the Research Data Assistance Center (ResDAC) at: http://www.resdac.umn.edu/

Medicare

1. **Research Identifiable Files (RIF)-** These Medicare and Medicaid files contain confidential beneficiary level data fields. Obtaining RIF data requires submission of a comprehensive information request package (See *Accessing CMS Research Identifiable Data)* to the Centers for Medicare and Medicaid Services (CMS). Following CMS approval for access to the RIF data, requests usually take 3 to 6 months to process. The cost of RIF data depends upon the file type and the size of the cohort and is determined by the processing time incurred by CMS. As an example, a request for one year of the Outpatient Standard Analytical File with a cohort size of less than one million would cost approximately $14,000. The Research Data Assistance Center at the University of Minnesota (http://www.resdac.umn.edu/) can provide data requesters with approximate cost figures.

    - **National Claims History (NCH) Nearline File-** This file is created from 100% of all institutional and non-institutional claims records submitted to CMS.

    - **Standard Analytical Files (SAF)-** These are final action claims level data files that are generated by the National Claims History (NCH) file and are available for the following types of providers:
        - Inpatient SAF
        - Skilled Nursing Facility SAF
        - Outpatient SAF
        - Home Health Agency SAF
        - Hospice SAF
        - Carrier SAF
        - Durable Medical Equipment SAF

    - **Medicare Provider Analysis and Review Files (MedPAR)-** Each record in the MedPAR file signifies a beneficiary stay in an inpatient hospital or a skilled nursing facility. Data fields summarize all services delivered from the time of admission to discharge. MedPAR data sets available include:
        - National
        - State
        - Skilled Nursing Facility

    - **Enrollment Database (EDB)-** Information in these data files includes beneficiary demographic and enrollment information. The following EDB's are available:

- Denominator File
- Vital Status File
- Cross Reference File
- Group Health Plan File (**Note: this file suspended as of April 2003 until HIPAA compliance issues resolved)**

- **Medicare Physician Identification and Eligibility Registry (MPIER)-** Data in this file include physician's unique identification number (UPIN), name, birth date, death date, medical school and year of graduation, sanctions, and practice information. (**Note: this file suspended as of April 2003 until HIPAA compliance issues resolved)**

- **Health Outcomes Survey (HOS) –** Information in these data files includes the entire national sample for a given cohort (including both respondents and non-respondents), and all of the HOS survey items. RIFs also contain protected beneficiary-level health information such as date of birth; however, specific direct person identifiers (i.e. name and health insurance claim number). The HOS RIFs include plan identifiers as well as several additional variables describing plan characteristics.

2. **Limited Data Set (LDS) Files-** Formerly known as Beneficiary Encrypted Files (BEF), many of these files duplicate RIF's but with certain data fields encrypted, ranged, or blanked to comply with HIPAA legislation. Additionally, there are less restrictive rules governing access (See *Accessing Limited Data Set Files)*. Following CMS approval for access to the LDS data, requests usually take 4 to 6 weeks to process. The costs of some LDS data are fixed, while others depend upon the file type and the size of the cohort and are determined by the processing time incurred by CMS.

    - **National Claims History (NCH) Nearline File-** This file is created from 100% of all institutional and non-institutional claims records submitted to CMS.

    - **Standard Analytical Files (SAF)-** These are final action claims level data files that are generated by the National Claims History (NCH) file and are available for the following types of providers:
        - Inpatient SAF
        - Skilled Nursing Facility SAF
        - Outpatient SAF
        - Home Health Agency SAF
        - Hospice SAF
        - Carrier SAF
        - Durable Medical Equipment SAF

- **Medicare Provider Analysis and Review Files (MedPAR)-** Each record in the MedPAR file signifies a beneficiary stay in an inpatient hospital or a skilled nursing facility. Data fields summarize all services delivered from the time of admission to discharge. MedPAR data sets available include:
    - National
    - State
    - Skilled Nursing Facility

- **Hospital Outpatient Prospective Payment System (OPPS)**- This file contains nearly 41 million records selected from claims level data for clinical outpatient services paid under the OPPS.

- **Medicare Current Beneficiary Survey (MCBS)-** This file contains MCBS interview data linked to Medicare administrative files to constitute a complete utilization data set. The goals of this survey are to determine expenditures and sources of payment for all services used by Medicare beneficiaries. The two data sets released from the survey are:
    - Access to Care
    - Cost and Use

- **Health Outcomes Survey (HOS) -** Information in these data files contain protected beneficiary-level health information such as date of birth; however, specific direct person identifiers (i.e. name and health insurance claim number) are not included in LDSs. The HOS LDSs include plan identifiers as well as several additional variables describing plan characteristics.

3. **Public Use Files (PUF)-** These files have been edited to remove any information that could be used to identify individual beneficiaries. In general the PUF's contain aggregate level information on Medicare providers.

    - **Medicare Cost Reports-** These files contains aggregate level information submitted by Fiscal Intermediaries to the Healthcare Cost Report Information Service (HCRIS). Cost reports available are:
        - Hospital
        - Skilled Nursing Facility
        - Renal
        - Hospice
        - Home Health

    - **Prospective Payment System (PPS) Files-** The information in these files is used to generate the PPS for institutional and non-institutional providers. The files available are:
        - SNF Wage Index
        - Hospital Wage Index Survey
        - Case Mix Index File

- ICD-9 Codes
- PPS Payment Impact File
- PPS Standardizing File
- Physician Fee Schedule Payment Amount
- Physician Fee Schedule Relative Value File
- Various other files

- **Physician Utilization and Demographic Information Files-** These files present summary statistics broadly describing physician utilization and demographics.
  - Unique Physician Identification Number (UPIN)
  - Medicare Disproportionate Share Adjustment
  - Top Diagnostic Related Groups (DRG) ranking
  - Top Current Procedural Terminology (CPT) ranking
  - Physician / Supplier Procedure Summary Master File

- **Hospital Survey, Address, and Utilization Files-** These files provide demographic data including zip code, length of stay, number of discharges, and total charges for hospitals and other provider organizations. Files include:
  - Provider of Services (POS)
  - Hospital Service Area

- **Health Outcomes Survey (HOS)** – The HOS PUFs contain the majority of the survey items collected on the HOS instrument (excluding beneficiary identifying information) as well as selected additional administrative variables. The files have been constructed in accordance with current CMS and Department of Health and Human Services (HHS) policies and other applicable statutes and laws. All identifying information has been excluded from the files, and demographic categories have been aggregated such that identification of any given individual is not possible.
  Three distinct categories of PUFs have been generated:
  - **Baseline PUFs** contain the data collected during a given baseline survey administration.
  - **Follow Up PUFs** contain the data collected during the follow up survey administration which occurs two years after the baseline survey administration.
  - **Analytic PUFs** contain the merged baseline and follow up files as well as supplemental variables.

Medicaid

1. **Research Identifiable Files (RIF)-** Similar to Medicare RIF data files that contain beneficiary level data, these require extensive documentation for access and use.

    - **State Medicaid Research Files (SMRF)-** SMRF files are calendar year files that report beneficiary enrollment, utilization, and expenditure data for Medicaid participants in 30 states.
      Variable Description:
      http://www.resdac.umn.edu/Medicaid/data_documentation.asp
      Additional Information:
      http://www.cms.hhs.gov/researchers/max/default.asp

    - **Medicaid Analytic Extract (MAX)-** Similar to SMRF but the MAX data files are available for all 50 states and include additional data fields.
      Variable Description:
      http://www.resdac.umn.edu/Medicaid/data_documentation.asp
      Additional Information:
      http://www.cms.hhs.gov/researchers/max/default.asp

2. **Public User Files (PUF)-** Numerous data files stripped of individual identifying data or presented in aggregate reports.

Additional Information:
http://www.cms.hhs.gov/medicaid/msis/mstats.asp
http://www.cms.hhs.gov/medicaid/mcaidsad.asp

Enrollment Database (EDB) Files

Overview: These demographic, enrollment, and entitlement files are the repositories of current and historic data collected for all beneficiaries ever enrolled in Medicare. The EDB is the primary source of demographic and geographic information for the entire Medicare program. EDB files are useful to researchers who wish to construct population samples and link to other various beneficiary level data files. Most EDB files are available as both Research Identifiable Files (RIF) and Limited Data Set (LDS) files. Access to EDB files is restricted and thus requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. For more information on accessing EDB data see *Accessing CMS Limited Data Set Files* or *Accessing CMS Research Identifiable Data.* All files are sent on CD-ROM, 3490e cartridges, or by special request on 3480 cartridges.

Data Available:

- **Denominator File-** This file combines Medicare beneficiary entitlement status information from the EDB with third-party payer information and Medicare managed care organization enrollment information. Unlike the broad EDB file, the Denominator file only contains data for beneficiaries who are enrolled during the given year.

- **Vital Status File-** This file contains demographic information for each beneficiary ever enrolled in Medicare. This file is frequently used to obtain death related information for a cohort of beneficiaries.

- **Cross Reference File-** This is not a freestanding file but is a creation of individual researcher data requests. When a researcher requests utilization or enrollment data from CMS they must specify their interest in obtaining the associated Health Insurance Claim (HIC) information. CMS constructs this HIC cross-reference file to identify and include all historical HIC's associated with each individual in the beneficiary cohort.

- **Group Health Plan File (Note: this file suspended as of April 2003 until HIPAA compliance issues resolved)-** This file contains information for all beneficiaries who have ever been enrolled in a managed care organization under contract with CMS. Each record represents a single beneficiary and includes demographic and managed care organization identification information.

Additional Information: For access to variable descriptions or downloadable data dictionaries see:

- Denominator File: http://www.resdac.umn.edu/ddde/dd_de.asp
- Vital Status File: http://www.resdac.umn.edu/docs/vtlstat.txt
- Cross Reference File: http://www.cms.hhs.gov/data/requests/xrefhic.pdf
- Group Health Plan File: http://www.resdac.umn.edu/docs/GHPMstr.txt

Medicare Current Beneficiary Survey (MCBS)

Overview: This is a continuous survey of a national sample of the Medicare population. The Office of Strategic Planning and CMS conduct the survey in order to provide a more complete picture of the use of health services, expenditures, and sources of payment for the Medicare population. The survey interviews selected beneficiaries three times a year to yield data that are designed to support cross-sectional and longitudinal analyses. Data from the MCBS are released in two files, the Access to Care file and the Cost and Use file. The Access to Care file is available faster than the Cost and Use file because it does not include utilization and expenditure related data. The Cost and Use file is more complete and includes complete information on beneficiary healthcare utilization and expenditure, including non-Medicare sources. There are two levels of identifiable information available in the MCBS data with the more sensitive including zip code level demographic information. MCBS data can be linked to Medicare claims and utilization data.

Access: Access to MCBS files is restricted and thus requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. For more information on accessing MCBS data see *Accessing CMS Limited Data Set Files.* All files are sent on CD-ROM, 3490e cartridges, or by special request on 3480 cartridges.

Survey Methods: A cohort of approximately 12,000 Medicare beneficiaries are surveyed every four months for a period of four years. The survey utilizes a nationally representative stratified cluster sample. The actual interview is administered utilizing computer-assisted interviews.

Additional Information:

- **Codebooks:** http://www.cms.hhs.gov/mcbs/Codebks.asp
- **Questionnaires:** http://www.cms.hhs.gov/mcbs/Quests.asp
- **SAS Input Statements:** http://www.resdac.umn.edu/MCBS/mcbs_sas_infile.asp
- **Data Documentation:** http://www.resdac.umn.edu/MCBS/data_documentation.asp

Medicare Health Outcomes Survey (HOS)

Overview: CMS, in collaboration with the National Committee for Quality Assurance (NCQA), launched the first Medicare managed care outcomes measure in the Health Plan Employer Data and Information Set (HEDIS®) in 1998. The measure includes the most recent advances in summarizing physical and mental health outcomes results and appropriate risk adjustment techniques. 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.

Access: Several types of Medicare HOS data files are available for research purposes. Medicare HOS data files are available as public use files (**PUFs**), limited data sets (**LDSs**), and research identifiable files (**RIFs**).

For more information go to: http://www.cms.hhs.gov/surveys/hos/hosdata.asp

HOS PUFs are available at no cost and can be downloaded directly from http://www.cms.hhs.gov/surveys/hos/hosdata.asp

A signed Data Use Agreement with CMS is required to obtain either LDS or RIF data files. The Research Data Assistance Center (ResDAC) at the University of Minnesota is a CMS contractor that provides assistance to academic, government and non-profit researchers interested in using Medicare and/or Medicaid data. ResDAC is available to review all request materials for Medicare HOS LDS and RIF data files for completeness and accuracy prior to their submission to CMS.

For additional information and assistance with obtaining Medicare HOS, LDS and RIF files, please visit the ResDac Medicare HOS Web page (www.resdac.umn.edu). ResDAC may also be contacted by calling 1-888-9RESDAC (1-888-973-7322) or by e-mail resdac@umn.edu.

Updated: Annually

Population: One thousand Medicare beneficiaries, who were continuously enrolled for a six month period, are randomly sampled from each plan and surveyed every spring (i.e., a survey is administered to a different baseline cohort, or group, each year). Two years later, these same respondents are surveyed again (i.e., follow up measurement). *Cohort I* was surveyed in 1998 and was resurveyed in 2000. *Cohort II* was surveyed in 1999 and was resurveyed in 2001. *Cohort III* was surveyed in 2000 and was resurveyed in 2002. *Cohort IV* was surveyed in 2001 and was resurveyed in 2003. During the current HOS administration in 2004, *Cohort VII* is being surveyed and *Cohort V* is being resurveyed.

Survey methods: The Medicare HOS instrument consists of three components: the SF-36® Health Survey; questions for case mix and risk adjustment purposes; and questions added by CMS as required by the 1997 Balanced Budget Act. Physical and mental functioning are measured with the Physical Component Summary (PCS) and Mental Component Summary (MCS) scores, which are derived from the SF-36®. For questions, contact the Medicare HOS Information and Technical Support Line. Telephone Line: 1-888-880-0077 E-mail Address: hos@azqio.sdps.org

Medicare Physician Identification and Eligibility Registry (MPIER)

(Note: this file suspended as of April 2003 until HIPAA compliance issues resolved)

Overview: This file contains information about authorized Medicare physician providers and their practice settings. Transamerica Occidental Life Insurance Company provides CMS with updated MPIER data each quarter. Each MPIER file relates to each active practice setting that a physician maintains. There are approximately 2.7 million records on file with quarterly updates adding 10,000 additional records. As a Research Identifiable File (RIF), MPIER file access is restricted and thus requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. For more information on accessing the MPIER data see *Accessing CMS Research Identifiable Data.*

Additional Information: For access to variable descriptions see:
http://www.resdac.umn.edu/docs/UPINM99.txt

Medicare Provider Analysis and Review Files (MedPAR)

Overview: MedPAR data are very similar to the data found in the Standard Analytical Files (SAF) except the unit of analysis for the SAF is a claim and for the MedPAR it is a stay (which may consist of several claims). By grouping data by stay the record summarizes all services delivered from admission to discharge during inpatient hospital and skilled nursing facility visits. This creates a fixed-length record for each hospital or skilled nursing facility stay, making it very attractive to researchers. Complete year files are available beginning in 1991 and the most current data are 3 years old with 100% file completeness. MedPAR's are available as both Research Identifiable Files (RIF) and Limited Data Set (LDS) files. Access to MedPAR data is restricted and information requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. All files are sent on 3490e cartridges or by special request on 3480 cartridges.

Data Available:

- **National-** This file contains 100% of the records for Medicare beneficiaries who used hospital inpatient facilities throughout the country during a calendar year. The cost is $3,655 per year.
- **State-** This file contains 100% of the records for Medicare beneficiaries who used hospital inpatient facilities in a particular state during a calendar year. The cost is $1,080 per state per year.
- **Skilled Nursing Facility-** This file contains 100% of the records for Medicare beneficiaries who used skilled nursing facilities throughout the country during a calendar year.

Additional Information: For access to variable descriptions and record layout see: http://www.resdac.umn.edu/ddme/dd_me_.asp

National Claims History (NCH) Nearline File

Overview: The NCH Nearline file is created from weekly claims records submitted to CMS from the nine CMS host sites. This file contains records of every raw institutional and non-institutional claim submitted, including all adjustment claims. These raw claims data include initial, interim, and debit/credit adjustments. The unit of analysis in this file is the claim. The file is divided into six record types. There are four institutional record types (Inpatient / SNF, Outpatient, Home Health, and Hospice) and two non-institutional record types (Physician / Supplier, and Durable Medical Equipment). Working with the NCH Nearline file can be cumbersome due to the number of adjustment claims present. The NCH file is available as both Research Identifiable Files (RIF) and Limited Data Set (LDS) files. Access to the NCH file is restricted and thus requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. For more information on accessing SAF data see *Accessing CMS Limited Data Set Files* or *Accessing CMS Research Identifiable Data.* All files are sent on 3490e cartridges or by special request on 3480 cartridges.

Additional Information: For access to variable descriptions, downloadable data dictionaries, and SAS input statements see:

- **Institutional files**: http://www.resdac.umn.edu/ddvib/dd_via2.asp
- **Non-Institutional files**: http://www.resdac.umn.edu/ddvib/dd_vib.asp

National Practitioner Data Bank (NPDB)

Overview: The NPDB acts as a clearinghouse of data relating to disclosable adverse events attributable to physicians, dentists, and other licensed health care professionals. Adverse events relate to malpractice payments, adverse licensure, clinical privileges, professional society membership, Drug Enforcement Agency (DEA) reports, and Medicare and Medicaid exclusion. Malpractice payers, state licensing agencies, hospitals, professional societies, and other healthcare related entities are required to report disclosable adverse events to the NPDB by law. NPDB information is used by healthcare providers in making clinical privilege and employment decisions. The NPDB data without information that identifies individual practitioners facilitates research into areas such as quality assurance, medical malpractice, licensing, and discipline.

Access: Data are available in ASCII or SPSS format from the NPDB website: http://www.npdb-hipdb.com/publicdata.html

For additional information contact:

Dr. Robert E. Oshel
Division of Practitioner Data Banks
Bureau of Health Professions
Health Resources and Services Administration
U.S. Department of Health and Human Services
7519 Standish Place, Suite 300
Rockville, MD 20857
Tel: 301-443-2300
Fax: 301-443-0238
E-mail: ROSHEL@HRSA.GOV

Cost: Free (Publicly Available)

Updated: Four times annually; March 31, June 30, September 30, and December 31

Population: Over 300,000 disclosable events among licensed healthcare providers

Survey Methods: Reporting entities submitted reports are consolidated into NPDB database format.

Linkage: No linkage available

Public Use Files

Medicare Cost Reports

Overview: These files contain aggregate level information submitted by Fiscal Intermediaries to the Healthcare Cost Report Information Service (HCRIS).

Cost reports available are:

- **Hospital**: http://www.cms.hhs.gov/data/download/hcris_hospital/default.asp
- **Skilled Nursing Facility:** http://www.cms.hhs.gov/data/download/hcris_snf/default.asp
- **Renal:** http://www.cms.hhs.gov/data/download/hcris_rnl/default.asp
- **Hospice:** http://www.cms.hhs.gov/data/download/hcris_hospice/default.asp
- **Home Health:** http://www.cms.hhs.gov/data/download/hcris_hha/default.asp

Prospective Payment System (PPS) Files

Overview: The information in these files is used to generate the PPS for institutional and non-institutional providers. The complete list of files available and their links can be obtained from: http://www.cms.hhs.gov/providers/pufdownload/default.asp#cmi

Physician Utilization and Demographic Information Files

Overview: These files present summary statistics broadly describing physician utilization and demographics.

Files available and links to information include:

- Unique Physician Identification Number (UPIN): http://www.cms.hhs.gov/data/alerts/upinalert.asp
- Top Current Procedural Terminology (CPT) Rankings, and Medicare Disproportionate Share Adjustment: http://www.cms.hhs.gov/statistics/feeforservice/default.asp
- Physician / Supplier Procedure Summary Master File: http://www.cms.hhs.gov/data/purchase/directory.asp#crspec

Hospital Survey, Address, and Utilization Files

Overview: These files provide demographic data including zip code, length of stay, number of discharges, and total charges for hospitals and other provider organizations. For more information regarding files available and links to information see: http://www.cms.hhs.gov/data/purchase/directory.asp - hospserv

Request for CMS Research Identifiable Data

Overview: Due to the confidential nature of Research Identifiable Files (RIF) requests for these data files involve extensive documentation and are subject to in-depth scrutiny by the Centers for Medicare and Medicaid Services (CMS). CMS no longer releases 100% files for RIF data. Data requests that are approved by CMS will only entitle the requesting researcher to those files that are relevant to the research project. Any alterations to the proposed research that necessitate obtaining additional RIF files requires the resubmission of a full request packet to CMS. The typical time frame for approval of requests for RIF data is between 3 and 6 months. Additional detailed information and access to the various forms is available on the CMS website. http://www.cms.hhs.gov/data/requests/academicandprivate.asp

As part of a request for RIF data the following documentation is required:

- **Written Letter of Request**- this letter outlines the purpose for the data request and includes a brief description of what data are required.

- **Data Use Agreement-** this legal document presents the requirements, policies, and procedures that the data requester is required to follow.

- **Data Request Form**- this simple form indicates which CMS data set is being requested and in what media format.

- **Privacy Board Review Summary Sheet**- Submission of this document indicates that the proposed research has IRB approval. This document must be completed by the researcher's Internal Review Board (IRB) and must be signed by an IRB board member.

- **Research Protocol**- this original document supplied by the requesting researcher provides detailed information relating to the proposed research. Elements that are required include: executive summary, objectives, background, importance, research question and methods, evaluation and analysis plan, work plan, staff and researcher qualifications, and implementation potential.

- **Evidence of Funding**- this form provides proof to CMS that the requester has the funds necessary to complete the research proposal.

If the proposed research is being federally funded, the researcher will also need to submit a letter of support from their federal project officer.

Completed request packets should be submitted to:

Robyn Thomas, Director
Division of Quality Coordination and Data Distribution (DQCDD)
OIS/EDG/DQCDD N1-15-03
Centers for Medicare and Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850

Standard Analytical Files (SAF)

Overview: These claims level data files are generated by the National Claims History (NCH) file and contain final action claims data. The final action claims include information collected by Medicare to pay for health care services provided to Medicare beneficiaries. The SAF's were developed in response to the cumbersome nature of the multiple adjustment claims present on the National Claims History Nearline file. SAF's are calendar year files that are updated quarterly. Complete year files are available beginning in 1991 and the most current data are 18 months old with 98.8% file completeness. SAF's are available as both Research Identifiable Files (RIF) and Limited Data Set (LDS) files. Access to SAF's is restricted and thus requests must be submitted to the Centers for Medicare and Medicaid Services (CMS) to obtain the files. For more information on accessing SAF data see *Accessing CMS Limited Data Set Files* or *Accessing CMS Research Identifiable Data.* All files are sent on 3490e cartridges or by special request on 3480 cartridges.

Data Available:

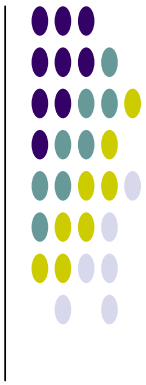
- **Inpatient SAF-** This file contains final action claims submitted by inpatient hospital providers for reimbursement. Data fields include diagnosis, procedure, diagnosis related group, dates of service, reimbursement amount, hospital, and beneficiary demographics.
- **Skilled Nursing SAF-** This file contains final action claims submitted by Skilled Nursing Facilities (SNF). Data fields are similar to those included in the Inpatient SAF, but also include the SNF provider number.
- **Outpatient SAF-** This file contains final action claims submitted by outpatient providers. Examples of outpatient providers include outpatient hospital departments (laboratory, radiology, ambulatory surgery, etc), clinics, renal dialysis centers, rehabilitation facilities, and mental health facilities. Data fields are similar to those included in the Inpatient SAF, but also include the outpatient provider number, and appropriate Healthcare Common Procedure Coding System (HCPCS) information.
- **Home Health Agency SAF-** This file contains final action claims submitted by home health agency providers. Data fields are similar to those included in the Inpatient SAF, but also include visit information including the number, type, dates, and provider number.
- **Hospice SAF-** This file contains final action claims submitted by hospice providers. Data fields are similar to those included in the Inpatient SAF, but also include the level of hospice care and the hospice provider number.
- **Carrier SAF-** This file which used to be known as the physician / supplier SAF contains final action claims data submitted by stand-alone providers (physicians, nurse practitioners, physician assistants, social workers, laboratories, and ambulance services). Data fields are similar to those included in the Inpatient SAF, but also include appropriate HCPCS and provider number.
- **Durable Medical Equipment (DME) SAF-** This file contains final action claims submitted by DME providers. Data fields are similar to those included in the Inpatient SAF, but also include the DME identifier number.

Additional Information: For access to variable descriptions, downloadable data dictionaries, and SAS input statements see:

- Inpatient, SNF, Outpatient, Home Health, and Hospice SAF files: http://www.resdac.umn.edu/ddvib/dd_via2.asp

- Carrier and Durable Medical Equipment SAF files: http://www.resdac.umn.edu/ddvib/dd_vib.asp

Public Data Sets



American Hospital Association (AHA) Annual Survey

Overview: The AHA Annual Survey data are used for research surrounding U.S. hospitals. This database of information is derived from the annual survey of member hospitals. Information provided includes more than 600 data fields covering Organizational Structure, Personnel, Hospital Facilities and Services, and Financial Performance.

Access: Information is in ASCII and Microsoft Access format on CD-ROM
Available through order from the AHA at:
http://www.ahaonlinestore.com/default.asp

For more information contact:
American Hospital Association
One North Franklin
Chicago, IL 60606
Tel: 800-242-2626
Fax: 312-422-4505
E-mail: storeservice@aha.org

Cost: $3,000 for current year; other years available at various prices

Updated: Annually since 1946

Population: Over 6200 hospitals and health systems

Survey Methods: Self-reported mail survey

Linkage: No linkage available

American Medical Association (AMA)

Overview: The AMA is the largest physician group in the United States. The AMA provides lobbying, continuing education, tools, and resources for the physician members. The AMA has a long history of collecting and maintaining data and is the single most comprehensive source for physician demographic data. The AMA collects information through surveys of physicians, medical groups, residency training sites, and medical schools. Additional information is obtained from various examination boards, accrediting organizations, and state and government agencies.

Data Collection Efforts

Census of Physicians

Collected since 1968, a mail questionnaire (PPA) has collected information on each physician's professional activity, practice specialty, type of practice, present employment, hospital affiliation, and group practice affiliation. These surveys are sent to the physician population at regular intervals.

Direct requests for PPA survey information to:

Susan Montrimas
Department of Census and Self Reported Data
515 N State St
Chicago, IL 60610
Tel: 312- 464-5184
Fax: 312- 464-4184
E-mail: Susan_Montrimas@ama-assn.org

Census of Medical Groups

The AMA periodically conducts the Census of Medical Groups (CMG) to monitor important trend in physicians' practice style. Recently the data collection methodology has changed from a paper survey every few years to telephone verification every six to nine months.

Direct requests for additional CMG information to:

Pat Scheibel
Department of Physician Practice and Communications Information
Tel: 312- 464-5432
Fax: 312 464-4184
E-mail: Pat_Scheibel@ama-assn.org

Annual Survey of Graduate Medical Education (GME) Programs

Approximately 7,900 accredited residency programs receive the Annual Survey of GME Programs. Program directors provide information on program characteristics such as clinical and research facilities and the learning environment. The Census of Graduate Medical Trainees portion of the survey collects data on individual residents in GME programs and is used to update physician AMA Physician Masterfile records.

Direct requests for additional GME survey information to:

Chris Mathews
Department of Census and Self Reported Data
Tel: 312- 464-4694
Fax: 312- 464-4880
E-mail: Chris_Mathews@ama-assn.org

Annual Survey of GME Teaching Institutions

Approximately 900 institutions sponsoring ACGME-accredited residency programs and 700 institutions that participate in GME training by making facilities available to one or more residency programs are surveyed annually. Respondents provide information on institutional features, resources, and benefits available to residents, and on institutional teaching staff.

Direct requests for additional GME survey information to:

Chris Mathews
Department of Census and Self Reported Data
Tel: 312- 464-4694
Fax: 312- 464-4880
E-mail: Chris_Mathews@ama-assn.org

Products Available:

AMA ePhysician Profile

This is an online database that provides vital information to individuals and organizations concerned with verifying physician credentials. Each profile includes primary source information on an individual physician's: name, birthday, degree, medical education, state licensure, certifications, federal sanctions, specialty, and others. The cost per physician profile is $29 or $27 for orders of three or larger. For access: https://profiles.ama-assn.org/amaprofiles/

American Medical Association Directory of Physicians in the United States

This annual directory is available as a four-volume text ($750) or a searchable CD-ROM ($750 single user; $1,500 2-4 users) with information on 820,000 physicians. The directory includes all U.S. physicians regardless of their AMA membership status. Information found in the directory includes the physician's name, mailing address, name of medical school and year of graduation, year first licensed, primary and secondary practice specialties, and American Board of Medical Specialties certification.

AMA Physician MasterFile

The MasterFile is the complete database from which the AMA's commercial products and databases are formed. The MasterFile is not available commercially, however the AMA will disseminate data to educational institutions, professional associations, and government agencies who are interested in collecting and analyzing physician data for ultimate dissemination. Each record includes the physician's name, medical school and

year of graduation, gender, birthplace, birth date, residency training, state licensure, board certification, geographical location, address, type of practice, present employment, and practice specialty. The MasterFile includes current and historical data on all physicians, living or deceased. Additional information is available regarding international medical graduates (IMG's). The data is drawn from a universe of 820,000 physicians, 19,000 medical practices, 7,900 accredited graduate medical education programs, 1,600 teaching institutions, and 125 accredited medical schools.

Requests for MasterFile data should include the purpose of the project, description of the specific information being requested, and the contact information of the requester.

Requests should be directed to:

Derek Smart
Department of Physician Practice and Communications Information
American Medical Association
515 N State St
Chicago, IL 60610
Tel: 312-464-4825
Fax: 312-464-4184
E-mail: Derek_Smart@ama-assn.org

Atlantic Information Services, Inc. (AIS)

Overview: AIS is a publishing and information company that develops data and strategic information for the health care industry. AIS's products are developed for providers, managed care plans, medical group practices, pharmaceutical companies, and other healthcare related professionals. AIS develops print and e-mail newsletters, loose-leafs, books, strategic reports, and databases. The databases available are:

- **Directory of Health Plans**- This database released annually provides a comprehensive listing of managed care organizations (MCO), preferred provider organizations (PPO), consumer directed health plans (CDH), and health plan enrollment by state. The data are grouped to cover national and state level enrollment by company, as well as national enrollment by type of product, and company demographic information. The database is available in print ($462) or as a CD-ROM ($1,495) stored Access or Excel database. For more information including data fields and examples go to: http://www.aishealth.com/Products/dhp.html

- **Employer Health Benefit Facts, Trends, and Data**- This data set was formed through a survey of employers to reveal health care benefit trends and strategies. Data includes employer utilization of managed care organizations, premium costs, and benefit structure. Each employer record indicates enrollment percentages for HMO, PPO, POS, and indemnity plans, percentages of fully insured versus self-insured, PBM's, and prescription drug benefit design. This survey is conducted periodically and is available as an Excel spreadsheet on CD-ROM ($339). For more information including data fields go to: http://www.aishealth.com/Products/efb.html

For further general information contact:

Atlantic Information Services, Inc.
1100 17th Street, NW
Suite 300
Washington, DC 20036
Tel: 800-521-4323
www.aishealth.com

Community Tracking Study (CTS)

Overview: CTS is a set of periodic surveys and site visits used to describe and analyze how the interactions of providers, insurers, policy makers and others determine the accessibility, cost, and quality of locally delivered health care. The study is a large-scale longitudinal investigation of health system change and its effects on people. It is designed to track a cohort of 60 American communities at two-year intervals. CTS collects data relating to households, employers, and physicians. The CTS is conducted by the Center for Studying Health System Change (CSHSC) and is sponsored by The Robert Wood Johnson Foundation.

Access: Public use files are accessible through the Health and Medical Care Archive (HMCA) and via CTSonline:
http://www.hschange.org/index.cgi?data=11
Restricted use data requires submission of a user agreement- for more info:
http://www.icpsr.umich.edu/HMCA/CTSform/forms.html

For general information contact:
Center for Studying Health System Change
600 Maryland Ave, SW #550
Washington, DC 20024
Tel: 202-484-5261
Fax: 202-484-9258
E-mail: ctshelp@hschange.org

Cost: Public use and restricted data files are free

Updated:

Household Survey	1996-97, 1998-99, 2000-01, 2003-04
Followback Survey	1996-97, 1998-99
Physician Survey	1996-97, 1998-99, 2000-01, 2004
Employer Survey	1997

Population:

Household Survey	Nationally representative survey of 32,000 families and 60,000 individuals
Followback Survey	28,000 household respondent's privately financed health insurance plans are followed back to the administering organization
Physician Survey	12,000 physicians spending at least 20 hrs/week in direct patient care
Employer Survey	22,000 public and private employers

Survey Methods: Primary method is by computer assisted telephone interview (CATI)
Households without telephones were provided with cellular telephones
Site visits to 12 communities are conducted every two years to interview health care leaders

Linkage: No linkage is available, including among separate CTS components

Dartmouth Atlas of Health Care

Overview: The Dartmouth Atlas project has produced a series of database resources in the form of annual and periodic national, regional, state, and condition-specific resources. The Atlas draws upon large databases including the 100% Medicare data, Blue Cross Blue Shield data, American Hospital Association data, and others to present the distinctive information. The Atlas is unique in that it includes a computerized geographical information system that describes the distribution and use of medical resources. The National Atlas is supplemented by nine regional sub-assessments to show the degree of variation in more than 40 population-based indicators. These indicators describe hospital expenditures and costs, hospital capacity, physician and specialist work forces, frequency of diagnostic and surgical procedures, and Medicare payments for care in 3,436 hospital service areas (HAS's) and 306 larger hospital referral regions in the United States. It also examines both the differences in per-capita health care resources (hospital beds, the physician work force, and spending) and the differences in how those resources are used (hospitalizations for common conditions, surgical procedures rates, and end-of-life care), depending on where patients live. The Dartmouth Atlas is conducted by the Center for the Evaluative Clinical Studies (CECS) under grant support from The Robert Wood Johnson Foundation (RWJF).

Access:

The data are available in a variety of formats including yearly data tables, comprehensive yearly Atlas', and regional supplements. The most recent Atlas is available by order. For access:
http://www.dartmouthatlas.com/download.php

There are also customizable interactive database tools including distribution graphs, benchmarking reports, community profile reports, and data tables. For access: http://geiger.dartmouth.edu/

For general information contact:

Dartmouth Atlas of Health Care
c/o Center for the Evaluative Clinical Sciences
Dartmouth Medical School
7251 Strasenburgh Hall
Hanover, NH 03755-3863
Tel: 603- 650-1820
Fax: 603- 650-1225

Cost:

All data except for the most current year's Atlas are free.
The most current year national Atlas and Data Viewer costs $1,295 and is available on CD-ROM and include data from every year the Atlas has been in existence. Current regional, state, and condition-specific Atlas's cost between $30 and $200.

Updated:

National Dartmouth Atlas- 1996, 1998, 1999 (data years)
Regional Atlas's- 1996 (data year)
State and condition-specific Atlas's- see website

Dorenfest

Summary: Sheldon I. Dorenfest & Associates, Ltd. provides management consulting services in the healthcare field and is now looked to as the market leader in providing information on information technology (IT) in the industry. Dorenfest & Associates collect the data for their database using telephone and mail surveys of CIO's and Vice Presidents of Marketing or Strategic Planning.

- **Integrated Healthcare Delivery System (IHDS+) + Database-** This database contains information from 1,500 integrated healthcare delivery systems consisting of 36,000 separate health care facilities. Each system is interviewed annually and updates are sent quarterly on CD-ROM. The database can be described by its two primary offerings. First, the demographic data section identifies the IHDS and its relationship with individual facilities. Information in the demographic data section includes:
    - Corporate ownership
    - Business strategy
    - Planned acquisitions
    - Key contact information
    - Staffed and licensed bed capacities
    - FTE's
    - Annual revenue and expenses
    - Number of operating rooms
    - Various other information

  Second, the information systems section contains current software and hardware data, as well as future IT purchase plans. Information included in the information systems section includes:
    - IT decision makers
    - IT strategy
    - IT budget and staff
    - Patient safety measures
    - HIPAA strategies
    - Computerized physician order entry (CPOE) applications
    - Picture archiving and communications system (PACS)
    - Software and hardware vendors with product and contract information
    - Purchase plans
    - Operating systems

  For a detailed sample view of the exact data available go to:
  http://www.dorenfest.com/SIXTHCOMPLETESAMPLEPROFILES.pdf

For additional general information contact:

Sheldon I. Dorenfest & Associates
515 N. State, Suite 1801
Chicago, IL 60610
Tel: 312-464-3000
www.dorenfest.com

Employer Health Benefits Survey

Overview: This survey jointly conducted by the Kaiser Family Foundation and The Health Research and Education Trust looks at employer sponsored health benefits. Information available includes the percentage of workers with job-based coverage, what kinds of plans employers are offering, what kinds of plans employees are choosing, and how much coverage costs everyone involved. Data are additionally segmented by industry, business size, region, and employee wage level. The survey excels at timeliness, with only an eight-month gap between data collection and release of results. The survey is widely used as an employer benchmarking tool to facilitate health benefit choices.

Access: Survey data are readily available in report form from the Kaiser Family Foundation Web site:
http://www.kff.org/insurance/index.cfm

Micro-data are released on a case-by-case basis.

Submission of a request for information should be sent to:

Sarah Williams
Kaiser Family Foundation
2400 Sand Hill Road
Menlo Park, CA 94025
Tel: 650- 854-9400
Fax: 650- 854-4800
E-mail: swilliams@kff.org

Cost: General reports and micro-data are available for free

Updated: Annually 1998 – Present

Population: 2000+ firms of varying size
Efforts are made to repeat sampling of firms from year to year

Survey Methods: The survey is a simple stratified probability sample. Computer assisted telephone interviews with human resource and benefits managers were performed by National Research LLC. Each company participating in the survey is asked as many as 400 questions about its largest conventional or indemnity, health maintenance organization (HMO), preferred provider organization (PPO), and point-of-service (POS) health plans.

Linkage: No linkage available

International City / County Management Association's (ICMA) Health Care Plans for Local Government Employees Data Set

Overview: ICMA is the professional and educational organization for local government managers, administrators, and assistants in cities, towns, counties, and regional entities throughout the world. The association provides a wide variety of services to its members and local governments including textbooks, development of professional ethics, newsletters, and survey data. ICMA conducts the Health Care Plans for Local Government Employees Survey, which was first conducted in 2002 with follow-up surveys planned for the future. The survey was mailed to 7,856 municipalities and counties. The survey asks respondents for information relating to type of health insurance, premiums and co-payments, eligibility, change in health insurance plans, cost switching measures, and others. For more information about the breakout of respondents' population size, geographic region, and metro status, or to look at the survey instrument see the web address below.

Access: The Excel database is available for purchase from the ICMA Bookstore at: http://bookstore.icma.org/obs/showdetl.cfm?&DID=7&Product_ID=1002&CATID=4

Cost: Academic/Public Sector - $550
Private Sector - $1,650

For more information contact:

International City/County Management Association (ICMA)
777 North Capitol Street, NE
Suite 500
Washington, DC 20002
Tel: 202-289-4262
Fax: 202-962-3500
Orders: 800-745-8780

InterStudy

Overview: InterStudy is a leading researcher and healthcare information retrieval service provider. They provide data, analysis, and directories of health maintenance organizations (HMO) and preferred provider organizations (PPO) information. InterStudy utilizes HMO and PPO primary survey data as the core of its analyses and databases. InterStudy's products include:

- **Managed Market Surveyor**- This product provides an overview of the managed care landscape at state, county, and metropolitan service area (MSA) levels. The product can be purchased and downloaded as either an Access or Excel file ($20,000). Information includes state, county, and MSA population, HMO and PPO enrollment, commercial HMO enrollment, market share and penetration, Medicare and Medicaid beneficiaries, uninsured and privately insured individuals, and plan level data. This product is updated annually since 1991. For further information including user's guide, methodology, and variable list go to: http://www.interstudypublications.com/ISHMOPPOManagedMarketSurveyor.asp

- **HMO Competitive Edge**- This series of databases provides HMO trend, market analysis, and enrollment information. The four databases available as Access or Excel files are:
  - **HMO Database**- Updated twice a year, this database provides information to compare HMO's in different markets. Information includes 3 years of enrollment data by product. ($2,400)
  - **Regional Market Analysis Database**- Updated twice a year, this database provides HMO information at an MSA level and includes market structure and competition analysis. ($2,300)
  - **Pharmacy Benefit Database**- This annually updated database draws upon nationally representative samples of information from national managed care firms, Blue Cross / Blue Shield plans, and HMO's to identify information including total enrollees with pharmacy benefits, total pharmacy expense by generic and branded prescriptions, and tiered co-payment schedules. ($7,000-$25,000)
  - **Contact Database**- Updated twice a year, this database provides key contact information for over 500 HMO's. (included in HMO Database)

  For further information including user's guide, methodology, and variable list go to: http://www.interstudypublications.com/ISHMOCompetitiveEdge.asp

- **HMO Financial**- This database details financial performance of publicly traded and privately held HMO's through regulatory agency disclosures. Available as Access or Excel files ($11,000 year end) these data reveal profitability, expenses, growth, and products offered. For more information go to: http://www.interstudypublications.com/ISHMOFinancial.asp

- **PPO Database**- This annually updated database presents data and analysis on the PPO industry. Reports can be customized by PPO, covered lives, and physician networks and generate information about size, rank, and physician participation at state and national

levels. The database is available as an Access or Excel file ($4,800). For further information including user's guide, methodology, and variable list go to: http://www.interstudypublications.com/ISPPODatabase.asp

- **PPO County Estimator**- This database provides the only source for county level PPO market share data. Additional information includes county level population and enrollment data. The database is available as an Access or Excel file ($9,200). For further information including user's guide, methodology, and variable list go to: http://www.interstudypublications.com/ISPPOCountyEstimator.asp

- **PPO MSA Estimator**- This database provides PPO data at the MSA level for each MSA in the country. Information includes population statistics, enrollment, and market share. The database is available as an Access or Excel file ($7,500). For further information including user's guide, methodology, and variable list go to: http://www.interstudypublications.com/ISPPOMSAEstimator.asp

For additional general information contact:

InterStudy Publications
2610 University Avenue West
Suite 500
St. Paul, MN 55114
Tel: 800-844-3351
www.interstudypublications.com

Joint Commission on Accreditation of Healthcare Organizations (JCAHO)

Overview: JCAHO is an independent, not-for-profit organization that evaluates and accredits more than 16,000 healthcare organizations and programs in the United States. JCAHO develops industry standards and then evaluates compliance of health care organizations in meeting these benchmarks. JCAHO makes limited organization accreditation information available for no charge from its website: www.jcaho.org Additional information is available for purchase from the JCAHO DataMart. The two DataMart products categories available consist of accreditation and demographic data and performance report data.

- **Accreditation and Demographic Data**- This online product allows users who have purchased a subscription to generate and retrieve customizable reports. Report outputs are available as ASCII comma-delimited, ASCII fixed field, Excel spreadsheet, or HTML page. Report fields include organizational identification and demographic information, accreditation decision, decision date, and survey due date. One week and one year subscriptions are available by organization type (ambulatory, behavioral, home care, hospital, laboratory, long term care, network, or all) with cost varying from $350 to $10,000.

- **Performance Report Data**- This online product allows users who have purchased a subscription to generate and retrieve customizable reports. Report outputs are available as ASCII comma-delimited, ASCII fixed field, Excel spreadsheet, or HTML page. Report fields include all those found in the Accreditation and Demographic Data product, as well as organizational scores in up to 46 performance areas with comparative data of other accredited organizations. One week and one year subscriptions are available by organization type (ambulatory, behavioral, home care, hospital, laboratory, long term care, or all) with cost varying from $1,000 to $15,000.

  For more complete data field descriptions and ordering information go to:
  http://dsa.trihost.com/store/default.asp?

For additional general information contact:

Joint Commission on Accreditation of Healthcare Organizations
One Renaissance Blvd.
Oakbrook Terrace, IL 60181
Tel: 630-792-5000
www.jcaho.org

Kaiser Data Request Form

Please complete the information below and return to bfinder@kff.org

Date:

Requested by:

Organization:

Address:

Phone:

Fax:

Email:

Please specify which document(s) you are requesting:

_____

Please specify your intended use of the data, below:

Medical Group Management Association (MGMA)

Overview: MGMA is the largest association for medical group practices with over 19,000 members managing 11,500 organizations and 237,000 physicians. Through its annual surveys of practice cost, physician compensation, and management compensation, MGMA produces benchmark data on medical practices. For access to all of the products, tables of contents, and current and archived survey questionnaires see: http://www.mgma.com/surveys/

For general information contact:

MGMA Headquarters
104 Inverness Terrace East
Englewood, CO 80112-5306
Tel: 303-799-1111; Toll-free: 877-275-6462
Fax: 303-643-4439

MGMA's products include:

- **Academic Practice Compensation and Production Survey for Faculty & Management**- This product contains data on more than 7,200 medical practice professionals and is useful for making financial and operational decisions in academic physician practice settings. Information includes compensation and productivity data for academic faculty by specialty and rank and other salary benchmarking data for administration. Information is based on the responses from 426 Clinical science departments. $495 (text)

- **Ambulatory Surgery Center Performance Survey**- This product contains data on more than 100 ambulatory survey centers operating throughout the U.S.. Information includes accounts receivable data, and cost data by square foot, per 1000 cases, per case, per procedure, per operating room, and per total revenue. Information is based on the responses from 113 ambulatory surgery centers. $340 (text)

- **Cost Survey**- This paper or CD-ROM product contains data on more than 1,200 medical organizations. Each of MGMA's Cost Surveys contain information on revenue and expenses, operating costs, staffing data, accounts receivable, charges, payer mix, and procedure volume. Information is based on the responses of 1231 medical organizations. $460 (text), $515 (CD-ROM), $875 (text & CD-ROM)

- **Cost Survey for Cardiovascular/Thoracic Surgery and Cardiology Practices**- This product contains information tailored for heart related practices. Data are drawn from over 100 cardiology practices and 30 cardiovascular/thoracic surgery practices. $460 (text)

- **Cost Survey for Integrated Delivery System Practices**- This product contains information on more than 380 hospital or integrated delivery system practices. $460 (text)

- **Cost Survey for Orthopedic Practices**- This product contains information on more than 130 orthopedic practices. $460 (text)

- **Performance and Practices of Successful Medical Groups**- This product contains data on more than 1,200 better performing medical organizations. This benchmarking tool allows comparison around costs, revenue, productivity, staffing, accounts receivable, and managed care indicators. Examples of best practices and recent articles from experts in the field are included. Information is based on the responses of 543 medical groups. $485 (text)

- **Physician Compensation and Production Survey**- This product contains data on more than 39,000 physician and non-physician providers working in more than 1,700 medical organizations. There are numerous data categories ranging from geographic region and practice setting to years in specialty and method of compensation. $475 (text), $515 (CD-ROM), $890 (text & CD-ROM)

- **Management Services Organization Performance Survey**- This product contains data on more than 70 management service organizations of various specialty, model, ownership, and size. Key performance metrics are included for benchmarking purposes. Information is based on the responses of 1434 managed services organizations. $460 (text)

- **MGMA Coding Profile Sourcebook: Primary Care Specialties**- This product describes data submitted by 1,143 medical practices with more than 50 million procedures performed by eight primary care specialties. Each of the MGMA Coding Profile Sourcebooks include coding data that makes benchmarking and trend analysis possible. $635 (text)

- **MGMA Coding Profile Sourcebook: Medical Specialties**- This product describes more than 14 million procedures submitted by 19 medical specialties. $635 (text)

- **MGMA Coding Profile Sourcebook: Surgical Specialties, Pathology and Radiology**- This product describes more than 11 million procedures performed by 14 physician and non-physician providers. $635 (text), $1,600 (set of all 3 Sourcebooks)

Medical Records Institute

Overview: The Medical Records Institute (MRI) serves the healthcare industry by providing a forum for sharing knowledge, experience, and solutions around the areas of electronic health records, e-health, and related information technologies (IT). They conduct conferences, seminars, and surveys dealing with Health IT subjects. The MRI Annual Survey of Electronic Health Record (HER) Trends and Usage is a survey of healthcare providers that was first conducted in 1999. For a preview of the survey see: http://www.medrecinst.com/uploadedFiles/resources/survey/surveyOverview03.pdf

Areas covered by the survey include:

- Management, administrative, and clinical motivations driving the need for Electronic Health Record systems
- EHR applications and functions being implemented or planned
- IT platforms used to support EHR applications
- EHR configurations for different environments
- Data capture methods being employed
- Major barriers to EHRs and the user strategies to address them
- Data security concerns
- Major trends and market shifts
- Analysis by respondent roles
- Detailed results by market segment
    - Hospitals
    - Solo and small group practices
    - Medium and large group practices
    - Integrated health delivery systems
- Respondent selection of vendors
- Strategic observations

Access: Current and archived versions of the survey are available in bound text as a PDF download and are available for purchase from the MRI website: http://www.medrecinst.com/iebms/reg/reg_p1_form.aspx?oc=10&ct=PUBLICATIONS&eventid=5016

Cost:
2003 - $250
2002 - $150
2001 - $100 (PDF download)
1999-2000 – Call

For more information contact:

Medical Records Institute
425 Boylston Street, 4th Floor
Boston, MA 02116-3315
Tel: 617-964-3923
Fax: 617-964-3926
E-mail: info@medrecinst.com

National Health Information (NHI)

Overview: NHI, LLC is a provider of specialized healthcare information for healthcare executives on various topics. NHI's primary products are newsletters and special reports, however they do provide results from an annual national survey on capitation.

- **Capitation Survey**- Published by NHI for eight consecutive years, this survey is based on over 500 respondents to NHI's survey of subscribers. Data include per member per month (PMPM) rate benchmarks for nearly all specialties, as well as for primary care, hospital services, and other categories. The data are separated by commercial versus Medicare populations and are presented in three-year format to show trends. The survey also includes data on days and admits per 1000, and length of stay. The survey is available in print ($99) only.

For more information contact:

National Health Information
P.O. Box 15429
Atlanta, GA 30333-0429
Tel: 800-597-6300
www.nhionline.net

National Survey of Alcohol, Drug, and Mental Health Problems (NSADMHP)

Overview: This survey is a component of The Robert Wood Johnson Foundation's Health Tracking Initiative. Focusing on care and treatment for alcohol, drug, and mental health conditions, the survey re-interviewed respondents to the 1996-1997 CTS Household Survey and the 1997-1998 Followback Survey as conducted by the Center for Studying Health System Change. Topics covered by the questionnaire include demographics, health and daily activities, mental health, alcohol and illicit drug use, use of medications, health insurance coverage including coverage for mental health, access, utilization, and quality of behavioral health care, work, income, and wealth, and life difficulties.

Access: NSADMHP data are restricted from general dissemination, but may be accessed with submission and acceptance of a Restricted Data User Agreement. Link to agreement information: http://www.icpsr.umich.edu/HMCA/CTSform/HCC/contents.html

For additional information:

Erik Austin
Health and Medical Care Archive
ICPSR
P.O. Box 1248
Ann Arbor, MI 48106-1248
Fax: (734) 647-8700
E-mail: hmca@icpsr.umich.edu

Cost: Free

Updated: One time collection 1997-1998

Population: 30,375 adult respondents with over sampling of poor, and users of mental health services

Survey Methods: Multistage stratified probability sample

Linkage: By matching on common identification variables, the data can be linked to the Community Tracking Study (CTS) Household Component, 1996-1997, and Followback Survey, 1997-1998.

National Survey of America's Families (NSAF)

Overview: NSAF is an ongoing series of a national survey used to track the effects of recent federal policy changes decentralizing many social programs. Data gathered include; economic, health, and social characteristics of children and families, participation in government programs, employment, earnings and income, economic hardship, educational attainment, training, family structure, housing arrangements, health insurance coverage, access to and use of health services, health status, psychological well-being, participation in religious and volunteer activities, knowledge of social services, and attitudes about work, welfare, health care, and childbearing. The NSAF is part of The Urban Institute's Assessing the New Federalism project.

Access: NSAF data releases are available for public use via the Urban Institutes web site: http://www.urban.org/content/Research/NewFederalism/NSAF/PublicUseData/PubUse.htm

For more information contact:

Assessing the New Federalism project
The Urban Institute
2100 M. St. NW
Washington, DC 20037
Tel: 202- 261-5377
Fax: 202- 293-1918
E-mail: nsaf@ui.urban.org

Cost: Free under user agreement

Updated: 1997, 1999, 2002 and future updates expected

Population: 43,000 national households representing 100,000 people in 13 states

Survey Methods: There were two separate components to the survey. One was a random digit dialing survey of households with telephones. A supplemental area survey was performed for households without telephones. Interviews were conducted on the telephone by interviewers working in central interviewing facilities, using computer-assisted telephone interviewing (CATI) technology. In-person interviewers used cellular telephones to connect respondents in households without telephones to the interviewing centers for the CATI interview.

Linkage: Currently no linkage is available

Panel Survey of Income Dynamics (PSID)

Overview: The PSID is conducted by the Survey Research Center, Institute for Social Research, at the University of Michigan and is funded through a grant from the National Science Foundation. PSID is an ongoing data collection effort begun in 1968 in an attempt to fill the need for a better understanding of the determinants of family income and its changes. The PSID has continued to trace individuals from the original national sample of approximately 4,800 households. One purpose for the study is that investigators hoped to discover whether most short-term changes in economic status are due to forces outside the family or if they can be traced to something in the individual's own background or in the pattern of thinking and behavior. The data can shed light on what causes family income to rise above or fall below the poverty line. Health related data include general health status, health expenditures, and insurance status.

Access: All PSID data are available for public use via the Internet at the PSID Web site: http://psidonline.isr.umich.edu/data/

For additional information:

The Panel Study of Income Dynamics
Institute for Social Research
PO Box 1248
Ann Arbor, MI 48106-1248
E-mail: psidhelp@isr.umich.edu

Cost: Free

Updated: Annually until 1997 and then biannually

Population: 6,168 families

Survey Methods: The PSID uses a cross-sectional national sample and a national sample of low-income families. The cross-sectional sample was drawn by the Survey Research Center (SRC). This was an equal probability sample of households from the 48 contiguous states. Between 1968 and 1972 the PSID was collected in face-to-face interviews using paper and pencil questionnaires. Thereafter, the majority of interviews were conducted over the telephone. In 1993, the PSID introduced the use of computer assisted telephone interviewing (CATI).

Linkage: None

Press Ganey

Overview: Press Ganey Associates is a large provider of patient and employee satisfaction measurement tools and improvement services for the health care industry. The company processes 8 million satisfaction surveys annually from 6,000 healthcare facilities in all 50 states, which represents 30% of the total acute care market. Press Ganey does not make available subsets of data collected from the various surveys, but it does make aggregate summary data reports available. These summary reports can highlight industry trends. For access to the findings reports see:
http://www.pressganey.com/products_services/readings_findings/findings/default.php

For more information contact:

Press Ganey Associates Corporate Headquarters
404 Columbia Place
South Bend, Indiana 46601
Tel: 800-232-8032
Fax: 574-232-3485

Quality Compass

Overview: Quality Compass is a tool to help employers, consultants, health plans, and researchers use the Health Plan Employer Data and Information Set (HEDIS) and Consumer Assessment of Health Plans (CAHPS) data more effectively. The National Committee for Quality Assurance (NCQA) created Quality Compass, a national database of HEDIS and NCQA Accreditation information from hundreds of health plans. Quality Compass makes it possible for users to look at health plans side by side to see how they compare, and thus to make health care coverage decisions based on quality and value, rather than on provider network and price. Quality Compass also includes national and regional averages and benchmarks, which help to establish targets for improvement and put plan results into a more meaningful context.

Access: Quality Compass data are available by order through NCQA Web site: http://www.ncqa.org/Info/QualityCompass/index.htm
Data are available in a easy user interface CD-ROM version or as an electronic data file in either SAS or Excel

For additional information:

NCQA
2000 L Street, NW
Suite 500
Washington, DC 20036
Customer Support Tel: 888-275-7585
Tel: 202-955-3500
Fax: 202-955-3599
E-mail: Customersupport@ncqa.org

Cost: The Quality Compass is broken into six different files available for Purchase separately or in packages. Data prices vary by number of users. Prices below are quoted for single user (see Web site for additional pricing information)

File 1- Plan Specific Data: Clinical, Access & Utilization- $5,500
File 2- Plan Specific Data: Member Satisfaction- $2,750
File 3- Summary Statistics: Clinical, Access & Utilization- $2,750
File 4- Summary Statistics: Member Satisfaction- $2,750
File 5- Blinded Detail Data: Member Satisfaction- $5,500
File 6- Summary and Plan Specific Medicaid Data- $2,000
Package 1- Files 1, 2, 3, & 4- $11,000
Package 2- Files 1 & 2- $6,600
Package 3- Files 1 & 3- $6,600
Package 4- Files 2 & 4- $4,400
Package 5- Files 1, 2, 3, 4, & 5- $15,400

Updated: Annually since 1996

Population: 300+ health plans covering over 120 million people

Survey Methods: The Quality Compass is a synthesis of HEDIS and CAHPS data that are gathered from a combination of mailed-surveys, administrative, and medical-record data, and accreditation checklists.

Linkage: No linkage available

Sherlock Company

Overview: The Sherlock Company provides information and consulting services to health plans and their stakeholders. Sherlock provides valuation, due diligence, and financial research consulting. Sherlock also offers a monthly newsletter, *PULSE* that provides health plan managers with monthly data to assess plan performance. Finally, Sherlock produces the Sherlock Expense Evaluation Reports (SEER) to serve as benchmark data for health plan performance.

- **SEER**- This database includes separate universes of publicly traded companies, Blue Cross / Blue Shield plans, and provider sponsored plans. The database is published annually since 1998 and is provided in both hard copy and Excel spreadsheet. The database consists of two volumes, Volume 1 ($11,500) contains financial metrics, and Volume 2 ($2,500 after purchase of Vol. 1) contains operational metrics.

  Volume 1 is separated into 9 distinct data sections per universe including:
  - Plan background, organization, and conventions
  - Summary analyses including functional expenses by product
  - Functional expenses across products
  - Functional expenses of each product, per member per month
  - Functional expenses of each product, percent of revenue
  - Functional expenses of individual products and national accounts
  - Functional expenses of pharmacy, mental health, and subrogation
  - Information systems expenses, allocated by supported functional area
  - Characteristics of respondent plans

  Volume 2 is separated into two principal data sections
  - Operational metrics by functional area
    - Sales and marketing
    - Broker commissions
    - Enrollment, membership, billing
    - Customer services
    - Provider network management
    - Claims and encounter capture and adjudication
    - Information services
    - Finance
    - Human Resources
  - Staffing costs and ratios
    - Employee compensation
    - Staffing costs per member, per employee, by functional area
    - Staffing ratios

For additional SEER information including a complete listing of data available go to: http://www.sherlockco.com/seerbackground.htm

For additional general information contact:

Douglas B. Sherlock, CFA
Sherlock Company
P.O. Box 413
Gwynedd, PA 19436
Tel: 215-628-2289
www.sherlockco.com

Solucient

Overview: Solucient L.L.P. is a leading provider of healthcare information retrieval services. Formed by the merger of HCIA-Sachs and HBS International, Solucient provides information, tools, and resources to providers, payers, managed care organizations, employers, pharmaceutical companies, and consultants. Solucient provides existing products as well as customizable data products. Solucient's products are divided into two classes, solutions and publication. Driving Solucient's extensive product lineup is its collection of comprehensive databases. These databases include:

- **Projected Inpatient Database (PIDB)-** The PIDB is the largest comprehensive payer inpatient database in the healthcare industry. Updated quarterly, the database is comprised of more than 22.6 million discharges per year from 2,900 hospitals, and represents 77.5% of all discharges. The PIDB is extrapolated to be universally representative of short-term, general, and non-federal hospital discharges. The databases' universe is defined using the Medicare Provider Analysis and Review File (MEDPAR) and the National Hospital Discharge Survey (NHDS). The database also utilizes a projection methodology that controls for age, sex, bed service category, census region, bed size, and teaching status.

- **Hospital Drug Utilization Database (HDUD)-** The HDUD is a collection of data from 150 acute care hospitals, representing over 1.9 million discharges. The database includes data from the 150 hospitals contract hospitals by capturing and standardizing drug usage. The HDUD is statistically adjusted for patient severity and provides universal data representative of short-term, general, and non-federal hospital discharges.

- **Claims Data Warehouse (CDW)-** The CDW accrues commercial claims data from 150 different payer sources. This database includes data for more than 9 million unique people. The CDW presents medical and pharmaceutical claims for longitudinal patterns of care for over 2 million people annually. The database is representative of the universe of the commercially insured population.

- **Medicare Database**- Solucient is responsible for maintaining the MEDPAR data for the Centers for Medicare and Medicaid Services (CMS). Solucient also purchases the annual 100 percent Standard Analytical Files (SAF) from CMS.

- **Cost Report and Financial Databases-** These databases house information relating to CMS reimbursements, including cost and charge data specific to the level of the revenue center. These data are used to calculate use and financial performance measures such as average length of stay, capital costs per adjusted admission, and expense per adjusted admission.

- **State Databases**- These databases collect hospital inpatient discharge data at the state level. Information in the databases includes patient characteristics, clinical information, and financial information. Solucient also supplements its data with purchases of state level data from various state hospital associations and governments.

Solutions:

- **Provider**
  - **ACTION O-I™-** This Internet based software package is essentially a customizable comparative database, reporting financial and operational data from over 270 departments in more than 900 hospitals across the country.
  - **HealthView Plus™-** This consumer market research product is offered collaboratively with Scarborough Research. The product provides consumer healthcare insight through national and local market surveys. The product provides 56 statistically significant, discrete population segments and provides insight into about consumer health status, chronic disease, attitudes toward health care, actual behavior, and preferences.
  - **Market Share Analysis**- Solucient will provide reports of market share, distribution, and use of services by hospital, zip code, and county.
  - **Physician Analysis**- Solucient will provide reports of physician-to-physician comparisons of resource use and clinical quality at facility, service line, DRG, and patient levels.
  - **ProviderView™-** This database consists of 900 combined clinical operations and financial elements for nearly every acute care and long-term care facility in the country. Solucient uses the most current MEDPAR data combined with five years of historical Medicare cost reports to produce the information. This database allows benchmarking on a department level for clinical measures and operational costs. The database is updated weekly.

- **Payer**
  - **Data-driven Length of Stay Norms-** These length of stay (LOS) data files include information from a database of 11 million discharges. The data are presented at an ICD-9-CM code level.

- **Employer**
  - **Claims Data Warehouse (CDW)-** This database contains 200 million claims records from hundreds of data sources, including insurance companies, pharmacy benefit managers, and specialty programs. Enrollment data forms the denominator in support of population-based studies. Longitudinal studies going back as far as five or six years are possible with this database. Possible types of analysis include:
    - Benchmarks/normative data development
    - Analyses by drug, device, diagnosis, or procedure
    - Trend analyses/cost studies
    - Regional analyses
    - Cost-benefit studies
    - Use rate studies
    - Longitudinal studies
    - Price pattern analyses
    - Service-level analyses
    - Rate setting

- **Pharmaceutical**
  - **ACTracker™-** This online database allows for analysis of drug and medical device utilization across a database of 550 unique hospitals and 13 million patient discharges.

Publications:

- **Hospital Benchmarking Tools and Guides**
  - **The Comparative Performance of U.S. Hospitals: The Sourcebook**- This benchmarking information source details hospital finance and utilization data including revenue, expenses, profitability, liquidity, capital structure, pricing strategies, utilization, capacity, patient and payer mix, and productivity. Includes annual information on nearly every acute care hospital in the U.S. This publication is available in formats including softbound, CD-ROM, and ASCII.
  - **Profiles of U.S. Hospitals**- This benchmarking information source provides performance and comparative data on 6,000 acute care and specialty hospitals in the U.S. Data includes otherwise hard to find details including percentage of Medicare and Medicaid patients served, profitability, occupancy rates, outpatient revenue percentages, average length of stay, days in accounts receivable, and top five diagnostically related groups (DRG's). This publication is available in formats including softbound, CD-ROM, and ASCII.
  - **The DRG Handbook**- This handbook includes detailed clinical, financial, and statistical data on over 100 of the most significant DRG's. Information includes comparisons across peer groups, average charge, average cost, average reimbursement, payer information, demographic breakdowns, and managed care penetration levels. This publication is available in formats including softbound, CD-ROM, and ASCII.
  - **Length of Stay Series**- This publication breaks down LOS by ICD-9-CM code and five age groups in diagnosis and operations. The data are drawn from a database of all payers representing 19 million discharges. This publication is available in formats including softbound, CD-ROM, and ASCII.

Access and Further Information:

For pricing information for publication products and further product information go to: http://www.solucient.com