



MEDICAL INDUSTRY LEADERSHIP INSTITUTE

Summary of Selected Medical Industry Data Resources

August 2018

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

Table of Contents

<i>Administrative Insurance Data</i>	1
Accessing CMS Limited Data Set Files	1
Accessing CMS Research Identifiable Data	3
Ambulatory Surgical Center (ASC) Payment System	6
Data Available from Center for Medicare and Medicaid Services (CMS)	7
Dartmouth Atlas of Health Care	10
Denominator File – LDS	11
Denominator File – RIF	12
End Stage Renal Disease (ESRD) Prospective Payment System (PPS) – LDS	13
Enrollment Database (EDB) Files	14
Health Care Cost and Institute (HCCI)	15
Home Health Prospective Payment System (HH PPS) Limited Data Set (LDS)	16
Hospital Outpatient Prospective Payment System (OPPS)	18
Hospital Outpatient Prospective Payment System (OPPS) Partial Hospitalization Program LDS	19
Implantable Cardioverter Defibrillator (ICD) Implantation Limited Data Set	20
Inpatient Psychiatric Prospective Payment System (IPFPPS)	21
Kaiser State Health Facts	22
Long-Term Care Hospital (LTCH) PPS Expanded Modified MEDPAR	24
Long-Term Care Minimum Data Set (MDS)	25
Medicaid Analytic eXtract (MAX)	26
Medicaid Analytic eXtract (MAX) Chartbook	27
Medicare Current Beneficiary Survey (MCBS)	28
Medicare Data on Provider Practice and Specialty (MD-PPAS)	30
Medicaid Pharmacy Benefit Use and Reimbursement	31
Medicare Provider Analysis and Review Files (MedPAR)	32
Medicare Provider Utilization and Payment Data	33
MEDPAR Limited Data Set (LDS) – Hospital (National)	34
Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers & Systems	35
National Claims History (NCH) Nearline File	36
Physician Compare Datasets	37
Physician Shared Patient Patterns Data	38

Provider of Services (POS) Current Files	39
Provider Specific Data for Public Use.....	40
Skilled Nursing Facility (SNF) MEDPAR Limited Data Set (LDS)	41
Standard Analytical Files (SAF)	42
Standard Analytical Files – LDS.....	44
<i>Administrative Provider Data</i>	46
Healthcare Cost and Utilization Project (HCUP).....	46
National Hospital Discharge Survey (NHDS)	48
<i>Population Survey Data</i>	50
Area Health Resource File (AHRF)	50
Behavioral Risk Factor Surveillance System (BRFSS)	51
Center for Disease Control and Prevention (CDC) - Pregnancy Risk Assessment Monitoring System (PRAMS)	52
Current Population Survey (CPS).....	54
Employer Health Benefits Survey	55
Health and Medical Care Archive (HMCA)	56
Health and Retirement Survey (HRS).....	57
Longitudinal Studies of Aging (LSOA)	59
Medical Expenditure Panel Survey (MEPS)	61
National Health Interview Survey (NHIS).....	63
National Survey of America's Families (NSAF)	64
Panel Survey of Income Dynamics (PSID)	65
Safety Net Monitoring Initiative.....	66
State and Local Area Integrated Telephone Survey (SLAITS) Child Well-Being and Welfare Module.....	67
State and Local Area Integrated Telephone Survey (SLAITS)- Health Module.....	68
Survey of Income and Program Participation (SIPP)	69
The Commonwealth Fund	70
<i>Survey Insurance Data</i>	71
HealthLeaders – InterStudy.....	71
<i>Survey Provider Data</i>	72
American Hospital Association (AHA) Annual Survey	72
American Medical Association (AMA)	73
Integrated Healthcare Delivery System (IHDS+) Database	76

Healthcare Integrity and Protection Data Bank	77
International City / County Management Association’s (ICMA) Health Care Plans for Local Government Employees Data Set.....	78
Joint Commission on Accreditation of Healthcare Organizations (JCAHO)	79
Medical Group Management Association (MGMA)	80
Medicare Health Outcomes Survey (HOS)	82
Medicare Physician Identification and Eligibility Registry (MPIER).....	83
National Ambulatory Medical Care Survey (NAMCS).....	84
Consumer Assessment of Healthcare Providers and Systems (CAHPS) Database	85
National Home and Hospice Care Survey (NHHCS)	86
National Hospital Ambulatory Medical Care Survey (NHAMCS).....	87
National Nursing Home Survey (NNHS)	88
National Practitioner Data Bank (NPDB).....	89
National Survey of Ambulatory Surgery (NSAS).....	90
<i>Other Data</i>	<i>91</i>
Atlantic Information Services, Inc. (AIS)	91
Community Tracking Study (CTS)	92
HcPro (formerly National Health Information NHI).....	93
Health Resources and Services Administration – Geospatial Data Warehouse.....	94
National Association of Health Underwriters	95
National Cancer Institute – Surveillance Epidemiology and End Result (SEER)	96
Press Ganey	97
Quality Compass National Committee for Quality Assurance (NCQA).....	98
Solucient.....	99
United States Department of Labor – Bureau of Labor Statistics (BLS).....	102
World Health Organization (WHO).....	103

Administrative Insurance Data

Accessing CMS Limited Data Set Files

Overview: The Centers for Medicare & Medicaid Services (CMS) makes Limited Data Set (LDS) files available to researchers as allowed by federal laws and regulations as well as CMS policy. LDS files contain beneficiary-level health information and are considered identifiable files, but they do not contain specific direct identifiers as defined in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. All disclosures of LDS files require a signed LDS Data Use Agreement (DUA) between the Centers for Medicare & Medicaid Services (CMS) and the Requestor.

LDS requestors are required to provide a research purpose as part of their request. The research purpose must relate to projects that could improve the U.S. healthcare delivery system. This includes projects related to improving the quality of life for CMS beneficiaries or improving the administration of CMS programs, including payment related projects and the creation of analytical reports. In addition, these research projects must contribute to generalizable knowledge. Please note that, since CMS publishes comments in the Federal Register, if you are requesting MedPAR files solely for the purpose of commenting on rulemaking, that meets the generalizable knowledge requirement.

LDS New Requires

1. Complete a LDS Request Packet that consists of the following:
 - **LDS request letter**, with the following information:
 - Project title
 - Brief overview of the use of the data
 - A contact person's name, organization, and e-mail address
 - **LDS Worksheet**
 - Use the LDS Worksheet to specify the files you are requesting. The Worksheet will automatically calculate the cost for your request.
 - If reusing data files from another DUA for your organization, do not select those files on the LDS worksheet. However, please note there is a \$600 administrative fee to reuse data unless you are only re-using Medicare Current Beneficiary Survey (MCBS) data. The re-use fee for MCBS is \$200. This fee is waived if new data files are being requested.
 - **LDS DUA Form 0235L**
 - Section 4 is where to list your requested file names and data years, and whether you are requesting a 5% sample or 100% of the file. If there are not enough rows to list all the files you need, please create and attach a continuation document

listing the additional file names and data years and annotate this in one of the rows in Section 4 of [DUA Form 0235L](#).

- If reusing data files from another DUA for your organization, please also list these files in Section 4 of [DUA Form 0235L](#).
 - **LDS DUA Attachment A – Research Protocol**
 - **[LDS DUA Attachment B](#) only** if you are requesting:
 - Standard Analytical Files (SAF) for 2012 or later, and/or
 - The NPI/UPIN Crosswalk.
2. Draft an email with the subject “New DUA-LDS < Requestor’s Organization Name>” and attach your LDS Request Packet (including signed DUA forms). If reusing data files from another DUA for your organization, include the files and years being requested and the number of the DUA from which you are reusing in the body of the email. It is not permissible to reuse data from another organization’s DUA.
 3. Send the email to the following contact:
 - If a CMS Sponsored Program/Awardee, Federal Grantee: Your CMS Contact or Federal Project Officer for them to submit on your behalf
 - All other researchers: DataUseAgreement@cms.hhs.gov.
 4. If your request is approved, the DUA team will notify you about how to submit your payment via www.pay.gov and how to forward that information to our payment coordinator.

Data Available and Cost

A complete list of the [CMS LDS files](#) is available on the ResDAC website. Individual file prices can be found on the [CMS Price List for Research Files](#).

Additional detailed information

https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA_-_NewLDS.html

Accessing CMS Research Identifiable Data

Overview: The Centers for Medicare & Medicaid Services (CMS) makes identifiable data files (IDFs) available to certain stakeholders as allowed by federal laws and regulations as well as CMS policy. Research identifiable files (RIF) contain beneficiary level protected health information (PHI) that can be used for analysis in your research study. A "study" is defined as having a strong research design, which clearly states the objectives and the significance of the research being conducted and provides a credible, straightforward argument for the importance of the project. The study shall address the following areas: hypotheses/study issues, data limitations, analysis plan and analysis methods. The scope and subject matter of the study must assist CMS in monitoring, managing, and improving the Medicare and Medicaid program and the services provided to beneficiaries. CMS must balance the potential risk to beneficiary confidentiality with the probable benefits gained from the completed research. A requestor must demonstrate expertise and experience to conduct and complete his/her proposed study using RIF data files.

Requests for RIF data require a Data Use Agreement (DUA) and are reviewed by CMS's Privacy Board to ensure that the beneficiary's privacy is protected and the need for identifiable data is justified. The [Researchers](#) page on the CMS website provides the criteria for the release of these data.

Requested Materials

REQUEST MATERIAL	DOCUMENTS	TIPS
Evidence of Funding	Download	Download , Download , Download
IRB Common Rule and HIPAA Waiver Approval		
Part D Identifier Attestation	Download	
Research Methods of Grant Proposal		
RIF Cost Estimate/Invoice	Download	
RIF Data Use Agreement (DUA)	Download	Download
RIF DUA - Attachment B	Download	Download
RIF Executive Summary Template	Download	Download
RIF Request Letter Template for Amendment	Download	Download
RIF Request Letter Template for New Study	Download	Download
RIF Specifications Worksheet	Download	

REQUEST MATERIAL	DOCUMENTS	TIPS
RIF Study Protocol Template		Download
Signature Addendum to DUA	Download	Download
Update to Existing Data Use Agreement	Download	Download

Additional Supporting Materials (if applicable)

REQUEST MATERIAL	DOCUMENTS	TIPS
Amendment Letter of Support from Federal Project Officer	Download	Download
Evidence of CAHPS Approval		
Evidence of HEDIS Approval		
Evidence of HOS Approval		
Evidence of HRS Approval		
Evidence of NHATS Approval		
IRF-PAI Variable Justification Worksheet	Download	
MDS 2.0 Variable Justification Worksheet - Nursing Homes	Download	
MDS 2.0 Variable Justification Worksheet - Swing Beds	Download	
MDS 3.0 Variable Justification Worksheet - Nursing Homes & Swing Beds	Download	
NDI Executive Summary Supplement	Download	
OASIS Variable Justification Worksheet	Download	
Part D Variable Justification Worksheet		
Reuse Letter of Support from Original Federal Project Officer	Download	Download

REQUEST MATERIAL	DOCUMENTS	TIPS
Reuse Letter of Support from Original Federal Project Officer - NIA	Download	
Reuse Letter of Support from Original Requestor	Download	Download

Request Materials and Submission Information

Email all required request materials to resdac@umn.edu. For the initial submission, please send draft documents only (no signatures). Following ResDAC's review and your completion of appropriate revisions to the request materials, ResDAC will forward them to CMS on your behalf. Data requests are reviewed by CMS on a rolling basis.

If you have multiple requests, please submit them individually for accurate tracking in ResDAC's ticketing system. For example, instead of submitting one new use request, two amendment requests, and all request documents in a single email, please submit three separate emails with only the documents specific to each request attached.

Data Available and Cost

Review all [Research Identifiable Files \(RIF\)](#) available for request. If you would like to obtain a formal cost estimate, please fill out the [Cost Estimate Request form](#) and submit to resdac@umn.edu. To obtain ballpark pricing, please call the ResDAC Assistance Desk.

Additional detailed information

<https://www.resdac.org/cms-data/request/research-identifiable-files>

Ambulatory Surgical Center (ASC) Payment System

Overview: This file contains a summary of service utilization by ASC supplier and is derived from 2016 ASC line item level data, updated through March 2017, that is, line items for services furnished on or after January 1, 2016 through December 31, 2016 that were received, processed, paid, and passed to the National Claims History file by March 31, 2017. This file contains data for ASC services billed in 2016, including services that were added under the revised ASC payment system. We revised CY 2016 HCPCS codes (where appropriate) to reflect their CY 2018 payment status and description using internal CY 2016 to CY 2017, and CY 2017 to CY 2018 crosswalk files. This file includes 754,090 records summarized first at the supplier level and then at the HCPCS code level. This is a flat file that is available on DVD.

The ASC LDS file will be available twice a year, once for the NPRM in late spring/early summer, and again for the final rule within a month of publication of the OPSS final rule in the Federal Register.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: Final Rule 2018

Cost: \$1,600

Media: DVD

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/ASCPaymentSystem.html>

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

For additional comprehensive information on data available from CMS access the Research Data Assistance Center (ResDAC) at: <http://www.resdac.umn.edu/>

Medicare

- **Research Identifiable Files (RIF)**
 - [UPIN Member File](#)
 - [Beneficiary Summary File](#)
 - [Beneficiary Annual Summary File](#)
 - [Chronic Condition Summary File](#)
 - [Part D Denominator](#)
 - [Denominator RIF](#)
 - [Consumer Assessment of Healthcare Providers & Systems RIF](#)
 - [Renal Management Information System](#)
 - [Value Modifier Practice Level File](#)
 - [Value Modifier Beneficiary Level File](#)
 - [Part D Drug Characteristics File](#)
 - [Part D Prescriber Characteristics File](#)
 - [Part D Pharmacy Characteristics File](#)
 - [Part D Drug Event File](#)
 - [MPIER File](#)
 - [Vital Status File](#)
 - [National Health and Aging Trends Study \(NHATS\)- Medicare Linked Data](#)
 - [Part D Plan Election Type Beneficiary Summary File](#)
 - [Part D Medication Therapy Management Data File](#)
 - [Home Health Outcome and Assessment Information Set](#)
 - [Health and Retirement Survey- Medicare Linked Data](#)
 - [Medicare Encounter Data Files](#)
 - [Medicare Data on Provider Practice and Specialty \(MD-PPAS\)](#)
 - [Plan Characteristics File](#)
 - [Health Outcomes Survey RIF](#)
 - [Inpatient \(Fee-for-Service\)](#)
 - [MedPAR](#)
 - [Hospice \(Fee-for-Service\)](#)
 - [Home Health Agency \(Fee-for-Service\)](#)
 - [Outpatient \(Fee-for-Service\)](#)
 - [Skilled Nursing Facility \(Fee-for-Service\)](#)
 - [Durable Medical Equipment \(Fee-for-Service\)](#)
 - [Carrier \(Fee-for-Service\)](#)
 - [Part D Formulary File](#)
 - [Value Modifier NPI Practice Level File](#)
 - [Pioneer Accountable Care Organization \(ACO\)](#)
 - [Long Term Care Minimum Data Set \(MDS\) 3.0](#)

- [Long Term Care Minimum Data Set \(MDS\) 2.0](#)
- [Inpatient Rehab Facility-Patient Assessment Instrument](#)
- [Medicare-Medicaid Linked Enrollee Analytic Data Source](#)
- [Long Term Care Minimum Data Set \(MDS\)- Swing-Bed 3.0](#)
- [Long Term Care Minimum Data Set \(MDS\)- Swing-Bed 2.0](#)
- [Master Beneficiary Summary File](#)
- [Healthcare Effectiveness Data and Information Set RIF](#)
- [Medicare Shared Savings Program Accountable Care Organizations \(ACO\) Beneficiary-level RIF](#)
- [Medicare Shared Savings Program Accountable Care Organizations \(ACO\) Provider-level RIF](#)
- [Risk Score Files](#)
- **Limited Data Set (LDS) Files**
 - [Inpatient Psychiatric Facility Prospective Payment System](#)
 - [Supplemental Date File LDS](#)
 - [Implantable Cardioverter Defibrillator \(ICD\) Implantation LDS](#)
 - [Skilled Nursing Facility LDS](#)
 - [Outpatient LDS](#)
 - [Carrier LDS](#)
 - [Ambulatory Surgical Center Payment System File](#)
 - [Durable Medical Equipment LDS](#)
 - [ESRD Prospective Payment System](#)
 - [Home Health Agency LDS](#)
 - [Hospice LDS](#)
 - [Inpatient LDS](#)
 - [MedPAR Hospital National LDS](#)
 - [MedPAR Longterm Care Hospital PPS Expanded Modified](#)
 - [MedPAR Skilled Nursing Facility LDS](#)
 - [OPPS Partial Hospitalization Program](#)
 - [Outpatient Prospective Payment System](#)
 - [Master Beneficiary Summary File LDS](#)
 - [Denominator LDS](#)
 - [Historical MCBS Data Files](#)
 - [Health Outcomes Survey LDS](#)
 - [LDS SAF NPI/UPIN Crosswalk](#)
 - [Home Health-OASIS Limited Data Set \(LDS\)](#)
 - [Medicare Current Beneficiary Survey \(MCBS\)](#)
- **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.
 - [NPI Downloadable File](#)
 - [Healthcare Cost Report Information System](#)
 - [Prescription Drug Plan Formulary and Pharmacy Network Files](#)
 - [Provider of Services File](#)
 - [UPIN Group File](#)
 - [Medicare Compare](#)

- [Medicare Geographic Variation Files](#)
- [Data Entrepreneurs' Synthetic Public Use Files \(DE-SynPUF\)](#)
- [Consumer Assessment of Healthcare Providers & Systems PUF](#)
- [Physician Supplier Procedure Summary Master File](#)
- [Value Modifier Performance PUF](#)
- [Medicare-Medicaid Linked Enrollee Analytic Data Source \(MMLEADS\) PUF](#)
- [Hospital Service Area File](#)
- [Part B Summary Data Files](#)
- [UPIN Directory](#)
- [Medicare Basic Stand Alone Claims Files](#)
- [Health Care Information System \(HCIS\)](#)
- [Healthcare Effectiveness Data and Information Set PUF](#)
- [Health Outcomes Survey PUF](#)

Medicaid

- **Research Identifiable Files (RIF)** - Similar to Medicare RIF data files that contain beneficiary level data, these require extensive documentation for access and use.
 - [Home Health Outcome and Assessment Information Set](#)
 - [Medicaid Enrollee Supplemental File](#)
 - [MAX Inpatient File](#)
 - [MAX Long Term Care File](#)
 - [MAX Other Therapy File](#)
 - [MAX Personal Summary File](#)
 - [MAX Prescription Drug File](#)
 - [Long Term Care Minimum Data Set \(MDS\) 3.0](#)
 - [Long Term Care Minimum Data Set \(MDS\) 2.0](#)
 - [Medicare-Medicaid Linked Enrollee Analytic Data Source](#)
 - [Long Term Care Minimum Data Set \(MDS\)- Swing-Bed 3.0](#)
 - [Long Term Care Minimum Data Set \(MDS\)- Swing-Bed 2.0](#)
- **Limited Data Set (LDS) Files**
 - [Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers & Systems LDS](#)
- **Public User Files (PUF)**
 - [Medicare-Medicaid Linked Enrollee Analytic Data Source \(MMLEADS\) PUF](#)
 - [Medicaid State Drug Utilization File](#)
 - [MAX Provider Characteristics File](#)

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.org/tools/downloads.aspx>

The Atlas Rate Generator allows researchers to obtain cross-sectional rates of Medicare data efficiently, at a low cost, and in a timely manner. For access: https://ctpr.dartmouth.edu/rate_requests/

For general information contact:

Dartmouth Atlas of Health Care

c/o The Dartmouth Institute for Health Policy and Clinical Practice
35 Centerra Parkway
Lebanon, NH 03766
PH: (603) 653-0800
FAX: (603) 653-0820

Denominator File – LDS

Overview: The Denominator File combines Medicare beneficiary entitlement status information from administrative enrollment records with third-party payer information and GHP enrollment information. The Denominator File contains data on all Medicare beneficiaries enrolled and/or entitled in a given year. It is an abbreviated version of the Enrollment Data Base (EDB) (selected data elements). It does not contain data on all beneficiaries ever entitled to Medicare. The file contains data only for beneficiaries who were entitled during the year of the data. These data are available annually in May of the current year for the prior year.

Note that the Denominator LDS file will be replaced by the Master Beneficiary Summary File (MBSF) LDS file. For calendar year 2016 data, either file can be requested. Beginning with data for calendar year 2017, only the MBSF LDS will be released.

Access: Please see *Accessing CMS Limited Data Set Files*

Available: CY 2006-2016

Cost: 5% file, \$250 per year/\$150 per quarter
100% file, \$1000 per year/\$625 per quarter

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/DenominatorLDS.html>

Denominator File – RIF

Overview: The Denominator File contains demographic and enrollment information about each Medicare beneficiary during a calendar year. The information in the Denominator File is 'frozen' in March of the following calendar year. This file is not available after March of 2010. This file was incorporated into the Master Beneficiary Summary File, Beneficiary file segment.

Some of the information contained in this file includes:

- beneficiary unique identifier
- state and county codes, zip code
- date of birth
- date of death
- gender
- ethnicity/race
- monthly entitlement indicators

Access: Please see *Accessing CMS Research Identifiable Data*

Available: This file was discontinued as of 03/2010 and is no longer available for request. The information in this file is now available in the Master Beneficiary Summary file, Base (a/b/d) segment.

Additional detailed information

<https://www.resdac.org/cms-data/files/denominator-rif>

End Stage Renal Disease (ESRD) Prospective Payment System (PPS) – LDS

Overview: This file contains select claim level data and is derived from 2017 claims for renal dialysis center services furnished on or after January 1, 2017 through December 31, 2017. The file contains claims with dates of service January 1, 2017 through December 31, 2017 that were received, processed, and passed to the National Claims History file by December 31, 2017. This file includes 3,670,286 claims for services furnished by renal dialysis centers that were paid under the ESRD PPS. This is a flat file available on DVD. The record length is 6,393 and block size is 32,760.

Access: Please see *Accessing CMS Limited Data Set Files*

Available: Proposed Rule 2019

Cost: \$600

Media: DVD

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/EndStageRenalDiseaseSystemFile.html>

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*.

CMS has updated the Medicare enrollment source data for the Master Beneficiary Summary File (MBSF). As of March 2017, the MBSF includes Medicare enrollment information from the CMS Common Medicare Environment (CME) rather than the CMS Enrollment Database (EDB). Data from the two sources was nearly identical. The CME improves the identification of Medicare Part B enrollment and also allows for more timely release of the MBSF. The universe of beneficiaries in the CME versus the EDB version of the MBSF are only slightly different.

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.
- **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 detailed information

<https://www.resdac.org/cms-data/file-family/Denominator>

<https://www.resdac.org/cms-data/files/vsf-name-and-address>

Health Care Cost and Institute (HCCI)

Overview: HCCI maintains an administrative claims database which includes enrollment files and medical and pharmacy (Rx) claims for over 50 million individuals per year from Aetna, Humana, and United HealthGroup. The membership data includes individuals and claims data from all 50 states and DC. The medical and Rx claims include allowed amounts (actual amounts paid) to providers as well as unique provider identifiers and member IDs that can be tracked over time. This core database is used to create multiple datasets that can be licensed for educational and research purposes.

The research datasets are intended for research purposes. The scope of the research and permissible uses of the results will depend on the type of license. The educational dataset is intended for instructional purposes and may only be used by students. No HCCI datasets may be used for commercial purposes or proprietary research.

HCCI's data contributors are committed to fostering a better understanding of the true drivers of increasing health care costs among thought leaders, policymakers and other health care decision-makers, and the general public. They include:

- Aetna
- Humana
- Kaiser Permanente
- UnitedHealthcare

Access: Application required. Please see website for more information
<http://www.healthcostinstitute.org/access-data/accessing-hcci-data-2018/>

Availability: CY 2017

Cost: [2018 Data Access Pricing](#)

Additional detailed information

<http://www.healthcostinstitute.org/access-data/accessing-hcci-data-2018/>

Home Health Prospective Payment System (HH PPS) Limited Data Set (LDS)

Overview: The “Home Health Claims – OASIS” Limited Data Set (LDS) file contains information on the utilization of the Medicare Home Health (HH) benefit. The file is constructed so that each observation represents a particular home health episode in a given year. Observations are stripped of most data elements that will permit identification of beneficiaries. As described in the proposed rule that presents the 2019 Home Health Prospective Payment System Rate Update and Proposed CY 2020 Case-Mix Adjustment Methodology Refinements, CMS is proposing refinements to the HH payment system that would change the length of a home health episode from 60-day to 30-days. That proposal meets the requirements of section 51001(a) of the Bipartisan Budget Act of 2018 to use 30-day periods. As such, this data file was constructed by splitting the current 60-day home health episodes into two 30-day periods. Each observation in the file represents a single 30-day period. However, some variables on the file were drawn from the 60-day episode from which the 30-day period was constructed, as indicated in the variable description. Information in this file includes:

- Start and end dates of the 30-day periods and 60-day episodes
- Wage index value associated with each episode/period
- Information regarding the resource use of the episode/period
- Payment adjustors used for the episode/period
- HIPPS codes
- Case-mix weights
- Indicators for whether the episode/period receives a payment adjustment (LUPA, PEP, outlier)
- Actual and simulated payments for the episode/period
- Information on number and length of visits that occur during the episode/period
- Select information from the Outcome and Assessment Information Set (OASIS) that is used in the payment system.

The “Home Health Claims – OASIS” LDS file will be updated twice a year.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: CY 2017

Cost: \$1,200

Media: DVD

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/Home_Health_PPS_LDS.html

Hospital Outpatient Prospective Payment System (OPPS)

Overview: This file contains select claim level data and is derived from 2016 hospital outpatient PPS claims, updated through June 2017, that is, claims for services furnished on or after January 1, 2016 through December 31, 2016 that were received, processed, paid, and passed to the National Claims History file by June 30, 2017. This file includes data elements such as diagnosis codes, bill type, outlier payments, and service revenue payments. This file includes more than 119 million claims for services paid under the OPPS, including multiple and single claims. The record length is 11,803, blocksize is 32,760.

The OPPS LDS file will be available twice a year, once for the NPRM in early summer, and again for the final rule within a month of publication of the OPPS final rule in the Federal Register.

NOTICE: Please note that as of April 2009, the OPPS Identifiable Data Set (IDS) is now considered a Limited Data Set (LDS). The former identifiable version of the file is now the only version available. The record layout below lists the data elements included on the new LDS, which include the elements from prior year versions of the IDS. This 2016 file has every element that was contained in the 2007 LDS and the 2007 IDS. Requestors should follow the link for Instructions on How to Request LDS Data Files below to obtain this file.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: Final Rule 2018

Cost: \$3,000

Media: DVD

Data Format: CSV format with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/HospitalOPPS.html>

Hospital Outpatient Prospective Payment System (OPPS) Partial Hospitalization Program LDS

Overview: This file contains select claim level data and is derived from 2016 claims for partial hospitalization program (PHP) services, with dates of service from January 1, 2016 - December 31, 2016 that were received, processed, and paid through June 30, 2017. This file includes data elements such as diagnosis codes, bill type, outlier payments, and service revenue payments. This file includes 76,353 claims for PHP services furnished by hospitals and community mental health centers (CMHCs) paid under the OPPS. This is a flat file. The LDS record length is 9,997, blocksize is 32,760.

The OPPS PHP file will be available twice a year, once for the OPPS Proposed Rule in late spring/early summer, and again for the Final Rule within a month of publication of the OPPS Final Rule in the Federal Register.

NOTICE: Please note that as of April 2009, the OPPS Partial Hospitalization Program (PHP) Identifiable Data Set (IDS) is now considered a Limited Data Set (LDS). The former identifiable version of the file is now the only version available. The record layout below lists the data elements included on the new LDS, which are the same elements included in prior year versions of the IDS. This 2016 file has every element that was contained in the 2007 LDS and the 2007 IDS. Requestors should follow the link for Instructions on How to Request LDS Data Files below to obtain this file.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: Final Rule 2018

Cost: \$1,140

Media: DVD

Data Format: CSV format with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/HospitalOPPSPHPLDS.html>

Implantable Cardioverter Defibrillator (ICD) Implantation Limited Data Set

Overview: The Medicare national coverage determination (NCD) for implantable cardioverter defibrillators (ICDs) requires that providers implanting these devices for certain clinical indications enter the implant data in a clinical registry. CMS temporarily operated a registry to collect ICD implantation data from February 2005 to April 2006. The ICD Implantation Limited Data Set (LDS) contains information CMS collected while operating this registry.

The ICD Implantation LDS includes almost 45,000 records with each record representing a device implantation. Each record contains approximately 60 data fields which include information about patient demographics, clinical history, clinical characteristics, device information, and the admission, implant and discharge dates. All records have been stripped of any direct patient identifiers.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: This is a stand-alone file with data from February 2005 to April 2006.

Cost: \$600

Media: CD ROM

Data Format: Text

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/ICD.html>

Inpatient Psychiatric Prospective Payment System (IPFPPS)

Overview: This file contains case level data for inpatient psychiatric stays and is derived from the 2016 MEDPAR data file and the latest available provider specific file. The data available in this file are consistent with the data as described in the IPF PPS Update Notice published in August 2017. The records are stripped of most data elements that would permit identification of beneficiaries. This file includes approximately 0.4 million records for services paid under the inpatient psychiatric benefit.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: August 2017 (FY 2018) IPF PPS Update Notice

Cost: \$3,000

Media: DVD

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/InpatientPsychiatricPPS.html>

Kaiser State Health Facts

Overview: State Health Facts is a project of the Henry J. Kaiser Family Foundation and provides free, up-to-date, and easy-to-use health data for all 50 states, the District of Columbia, and the United States. In some cases, data are available for counties, territories, and other geographies. State Health Facts is comprised of more than 800 health indicators and provides users with the ability to map, rank, trend, and download data. Data come from a variety of public and private sources, including Kaiser Family Foundation reports, public websites, government surveys and reports, and private organizations. Data presented on State Health Facts are updated or added as new data become available; the update schedule varies from indicator to indicator. The availability of older or “archived” data is limited and varies by indicator. Topics include:

- Demographics and the Economy
- Health Costs & Budgets
- Health Coverage & Uninsured
- Health Insurance & Managed Care
- Health Reform
- Health Status
- HIV/AIDS
- Medicaid and SCHIP
- Medicare
- Minority Health
- Providers and Service Use
- Women’s Health

Custom State Reports is an interactive tool for compiling health-related data for a single state or multiple states into a single report. Users can choose from hundreds of data indicators, specially selected from the Foundation’s State Health Facts collection, to build interactive reports that focus on specific health conditions or issues in a state or region most relevant to their work.

The Custom State Reports tool walks you through a four-step process. Data are organized into categories, subcategories, and indicators. Expand categories in the first step to reveal subcategories. In the second step, expand the subcategories to explore the specific indicators that are available. We recommend you choose 10 or fewer indicators for optimal performance. In the third step, you can choose up to 50 states, the District of Columbia and the United States to compare in your report.

Once a custom report is created, there are several ways to interact with the data:

1. **Refine the data** to look at different data views (e.g. numbers or percents), timeframes, or in some cases, the data are broken down further into distributions (by gender, age, etc.).
2. Click and drag state columns to **reorder the report**.
3. **Add data or states** to the report by clicking the “Add States” or “Add Indicators” button at the bottom of the report page.
4. **View notes and sources** of the data using the “Notes” button on the report page or by downloading the data to see a streamlined view of the notes and sources for the indicators you have chosen.

5. **Open full indicator** to see the indicator in State Health Facts. As stated above, State Health Facts will provide you the ability to map, rank, trend, and download the data, when available.
6. **Download the report** to be able to further manipulate the data in Excel. Excel will give you the ability to transpose, filter, and further organize the data as you wish.
7. **Share** the report by clicking the “Email,” “Share,” or “Copy Link” button. The “Copy Link” button will provide the URL of the report so it can be opened exactly as you have built it at any time.

Additional detailed information

<https://www.kff.org/statedata/>

Long-Term Care Hospital (LTCH) PPS Expanded Modified MEDPAR

Overview: The Medicare Provider Analysis and Review (MEDPAR) file contains records for 100% of Medicare beneficiaries who received hospital inpatient services at a LTCH as defined in Sections 1886(d)(1)(B)(iv) and 1861(ccc) of the Social Security Act. The file contains same data elements found in the National Hospital MEDPAR Limited Data Set (LDS), including the billed Diagnosis-Related Group (DRG), the current DRG and the proposed/final DRG. The records are stripped of most data elements that will permit identification of beneficiaries. The six position Medicare billing number identifies the LTCH. The file is available to persons qualifying under the terms of the Routine Use Act as outlined in the December 24, 1984, "Federal Register" and amended by the July 2, 1985, Notice. This file consists of between approximately 140,000 and 160,000 records.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: FY 2004 – FY 2016

Cost: \$100

Media: DVD

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/LTCHPPSMEDPAR.html>

Long-Term Care Minimum Data Set (MDS)

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

- physical
- clinical
- psychological
- psycho-social functioning
- life care wishes

The items in the MDS give a multidimensional view of the patient's functional capacities and helps staff to identify health problems.

MDS data from CMS can typically be requested within three months of the target date. The version of the MDS changed from 2.0 to 3.0 on 10/01/2010. The 4th quarter of 2010 data will be version 3.0.

Access: Please see *Accessing CMS Research Identifiable Data*

Available: CY 2010-2016

Additional detailed information

<https://www.resdac.org/cms-data/files/mds-3.0>

Medicaid Analytic eXtract (MAX)

Overview: The Medicaid Analytic eXtract (MAX) data – formerly known as State Medicaid Research Files (SMRFs) – are a set of person-level data files on Medicaid eligibility, service utilization, and payments. The MAX data are extracted from the Medicaid Statistical Information System (MSIS). The MAX development process combines MSIS initial claims, interim claims, voids, and adjustments for a given service into this final action event. Unlike fiscal-based MSIS quarterly files, MAX are annual calendar year files.

- [MAX Prescription Drug File](#)
- [MAX Personal Summary File](#)
- [MAX Other Therapy File](#)
- [MAX Long Term Care File](#)
- [MAX Inpatient File](#)

Access: Publicly downloadable

Available: CY 1999 - 2014

Additional detailed information

<https://www.resdac.org/cms-data/file-family/Medicaid-Analytic-Extracts-MAX>

Medicaid Analytic eXtract (MAX) Chartbook

Overview: The Medicaid Analytic eXtract (MAX) Chartbooks are research tools and reference guides on Medicaid enrollees and their Medicaid experience in 2002, 2004 and 2008. Developed for state Medicaid directors, policymakers, researchers, and others interested in the Medicaid program, the chartbooks consists of illustrative graphs, descriptive text, and an extensive data appendix with summary national- and state-level information on enrollee demographic and eligibility characteristics, Medicaid service use, and Medicaid expenditures in 2002, 2004 and 2008. The chartbooks summarize the Medicaid program and the MAX data system. They also present figures and tables that reflect the Medicaid population in 2002, 2004 and 2008. For each of the three chartbooks, Chapter 2 profiles Medicaid enrollees and their Medicaid experience nationally, Chapter 3 presents state-level detail, and Chapters 4 through 6 provide supplementary information on special topic areas, including managed care, dual Medicare/Medicaid enrollees, and service use and expenditures by detailed type of service. An appendix contains source data tables for Chapters 3 through 6. A new chapter has been added to the latest chartbook that presents data for 2008, Chapter 7 on Medicaid waiver enrollment and utilization. There is also a chartbook for MAX 2008 with a focus on patient “encounter” records for services delivered through prepaid managed care plans. This chartbook describes the service utilization of Medicaid enrollees in managed care plans in MAX 2008 data. The chartbook extends the analysis of the previous MAX chartbooks, which focused on the service utilization of Medicaid enrollees covered on a fee-for-service basis. This chartbook also supplements recent issue briefs focusing on the quality and completeness of encounter data. It provides valuable information for CMS and researchers on the availability of and uses for encounter data in MAX data.

Access: Publicly downloadable

Availability: 2002, 2004, 2008

Additional detailed information

https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX_Chartbooks.html

Medicare Current Beneficiary Survey (MCBS)

Overview: The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Office of Enterprise Data and Analytics (OEDA) of the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The central goals of the MCBS is to determine expenditures and sources of payment for all services used by Medicare beneficiaries, including co-payments, deductibles, and non-covered services; to ascertain all types of health insurance coverage and relate coverage to sources of payment; and to trace outcomes over time, such as changes in health status and spending down to Medicaid eligibility and the impacts of Medicare program changes on satisfaction with care and usual source of care.

Current Availability:

- MCBS Survey File – 2015 Available
- MCBS Cost Supplement File – 2015 Available
 - MCBS Survey File

The MCBS Survey File contains survey collected data augmented with Administrative data to allow for analysis regarding the beneficiaries' health status, access to health care, satisfaction with health care and usual source of care. The following information is contained in the MCBS Survey File: Beneficiary Demographics, Household Characteristics, Access to Care, Satisfaction with Care, Usual Source of Care, Health Insurance Timeline (shows types of insurances, the coverage eligibility, and what is covered), Health Status and Functioning and other topical survey sections like Medical Conditions, Health Behaviors, Preventive Services, Interview Characteristics, Beneficiary Knowledge of the Medicare Program, Residence Timeline, Facility Characteristics, and Beneficiary Income and Assets. This file also includes summarized administrative FFS utilization data and research claims which contain limited FFS claims content for those beneficiaries enrolled in FFS Medicare and Medicare Part D. This file is released 12-15 months after the end of data collection.

- MCBS Cost Supplement File

The MCBS Cost Supplement File provides cost and utilization data that can be linked to the MCBS Survey File to conduct analysis on healthcare cost and utilization for the beneficiaries in the survey. This is not a stand-alone file. Users of the Cost Supplement File will now require the Survey File for information on beneficiaries' demographic characteristics and health insurance information, as these fields are no longer included in the Cost Supplement File. The MCBS Cost Supplement file links Medicare claims to survey-reported events and provides complete expenditure and source of payment data on all health care services, including those not covered by Medicare. The MCBS Cost Supplement file provides a comprehensive picture of health services received, amounts paid, and all sources of payment. The file can support a broader range of research and policy analyses on the Medicare population than would be possible using either survey data or administrative claims data alone. Survey-reported data include information on the use and cost of all types of medical services including those not covered by Medicare, as well as information on supplementary health insurance costs. The Cost Supplement file includes use and cost information on dental care, inpatient hospitalizations, outpatient hospital care, physician services, durable medical equipment, skilled nursing home and other facility services,

prescription medications, and other medical services. This file is released 15-18 months after final administrative and claims data are available.

For data files 1991-2013 there are two data files from the Medicare Current Beneficiary Survey (MCBS) that were released in annual Access to Care and Cost and Use files, which can be obtained directly from CMS. 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.

Cost: MCBS Access to Care: \$600 per year, 1991-2013 available
MCBS Cost and Use Files: \$600 per year, 1992-2013 available

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MCBS.html>

Codebooks: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks.html>

Questionnaires: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires.html>

Medicare Data on Provider Practice and Specialty (MD-PPAS)

Overview: The MD-PPAS file assigns Medicare providers to medical practices based on the tax identification numbers and elaborates on the Centers for Medicare & Medicaid Services (CMS) provider specialty classification. This provider-level dataset is built around two identifiers: the national provider identifier (NPI) and the tax identification number (TIN).

The MD-PPAS has been updated to version 2.3 from version 2.2 to reflect the following changes:

- The data now include variables on the percentage of a provider's line items that are performed in eight place-of-service categories.
- In 2017, PECOS introduced a new hospitalist specialty designation. To accommodate this change, version 2.3 changes the definition of hospitalists to include any provider who self-identifies as such in the PECOS data as well as primary care physicians who had at least 90% of their line items in an inpatient hospital setting.
- The data source for the part B non-institutional claims data changed the method by which it identifies final action claims. This modification can change the content of the claims data, e.g. procedure codes and allowed-charges dollar amounts. Consequently, there are some providers who no longer satisfy the criteria for inclusion in the MD-PPAS, leading the provider counts in the latest version of MD-PPAS to decline slightly.
- The data use a corrected form of the zip-code-to-CBSA crosswalk to assign providers to CBSA based on the zip code reported on their submitted claims. This correction resulted in a small percentage of providers that exist in both version 2.2 and version 2.3 with different values for their CBSA code and type variable.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: FY 2008 – 2016

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

<https://www.resdac.org/cms-data/files/md-ppas>

Medicaid Pharmacy Benefit Use and Reimbursement

Overview: Under a research contract from CMS’s former Office of Research, Development and Information (ORDI), now the Office of Information Products and Data Analytics, Mathematica Policy Research, Inc. (MPR) is producing a series of research products related to pharmacy benefit use and reimbursement in Medicaid. MPR is using the Medicaid Analytic eXtract (MAX) data files for calendar year 1999 and later years for these research products. The MAX files are constructed from claims and eligibility data that states submit to CMS electronically through the Medicaid Statistical Information System (MSIS).

One set of products is the Statistical Compendium: Medicaid Pharmacy Benefit Use and Reimbursement (hereafter “the Compendium”) for calendar years 1999 and 2001-2009. The Compendium for each year provides detailed state-by-state and national data on the use of and reimbursement for prescription drugs in Medicaid. (A Compendium was not produced for calendar year 2000.)

Using the detailed data from the Compendium, MPR has also prepared Chartbooks for 1999 and 2001-2009. The Chartbooks present selected highlights from the Compendium and include comparisons across states. The 2001-2009 Chartbooks also include comparisons to earlier years.

Access: Publicly downloadable

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MedicaidPharmacy.html>

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).

MedPAR Files contain inpatient hospital and/or skilled nursing facility (SNF) final action stay records for all Medicare beneficiaries. MedPAR files contain the following information:

- procedures, diagnoses, and DRGs
- length of stay
- beneficiary and Medicare payment amounts
- summarized revenue center charge amounts

Each MedPAR record may represent one claim or multiple claims, depending on the length of a beneficiary's stay and the amount of services used throughout the stay. 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.

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.

Available: CY 1999 – 2016
Next year request: December 2018

Additional Information: For access to variable descriptions and record layout see:
<https://www.resdac.org/cms-data/file-family/MedPAR>

Medicare Provider Utilization and Payment Data

Overview: CMS has released a series of publicly available data files that summarize the utilization and payments for procedures, services, and prescription drugs provided to Medicare beneficiaries by specific inpatient and outpatient hospitals, physicians, and other suppliers. These Medicare Provider Utilization and Payment Data files include information for common inpatient and outpatient services, all physician and other supplier procedures and services, and all Part D prescriptions. Providers determine what they will charge for items, services, and procedures provided to patients and these charges are the amount that providers bill for an item, service, or procedure.

Access: Publicly downloadable

Data Format: Data are being made available in Microsoft Excel (.xlsx) format and raw text file data format (comma separated values (.csv) for inpatient and outpatient and tab delimited for physician and other supplier and Part D prescriber)

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html>

MEDPAR Limited Data Set (LDS) – Hospital (National)

Overview: MedPAR consolidates Inpatient Hospital or Skilled Nursing Facility (SNF) claims data from the National Claims History (NCH) files into stay level records. The accumulation of claims submitted for the period commencing on a beneficiary's date of admission to an inpatient hospital or SNF and ending on the beneficiary's date of discharge from that hospital or SNF represents one stay. In the case of a SNF stay where the beneficiary has not yet been discharged and remains a patient, the claims submitted between the admission date to the SNF through the time of the MedPAR file creation, represent one stay. A stay record may represent one or more final action claims.

NOTE: Any given MedPAR file represents a static snapshot of a specific stay at the time the data was sourced from NCH. As such, any given stay record on a specific update of MedPAR DOES NOT NECESSARILY represent the final coding and/or payment information for that stay because if subsequent adjustments to the claims that comprise the stay occur after MedPAR is run, they will not be reflected on the file.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: FY 1997 – FY 2016

Note: Regarding the 1999 File – The file contains corrupted data in data element "Admission of Date of Death Interval" positions 528-532. Corruption occurred during the Y2K conversion of legacy data files. At present, there are no plans to correct the data.

Cost: \$3,600 per year

Media: DVD

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MEDPARLDSHospitalNational.html>

Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers & Systems

Overview: The 2014-2015 Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers & Systems (NAM CAHPS) survey contains national and state-level estimates of the health care experiences of non-institutionalized adults (ages 18+) who were enrolled in Medicaid between October 2013 and December 2013. The survey used a modified version of the Adult CAHPS® Medicaid 5.0H survey and included supplemental questions to assess potential access barriers more broadly and capture data on race, ethnicity, sex, primary language, and disability status per Section 4302 of the Affordable Care Act. The sample population was stratified into four subgroups:

- Dually-eligible individuals
- Individuals with a disability (non-duals)
- Medicaid managed care (non-duals, non-disabled)
- Fee-for-service (traditional) Medicaid (non-duals, non-disabled)

The survey was conducted by CMS in four waves starting in the fall of 2014 and was based on a nationwide sample of roughly 29,000 adult Medicaid enrollees from each of the 46 participating states and the District of Columbia. This file does not contain any personally identifiable information (PII). The survey data can be used to produce measures of:

- Demographic and health characteristics
- Access to care
- Barriers to care
- Global ratings of care

Access: Please see *Accessing CMS Limited Data Set Files*

Available: FY 2014 (Oct 2013 – Dec 2015)

Next Year Request: To be determined based on availability of funding.

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/CAHPS.html>

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*.

Additional detailed information

<https://www.resdac.org/cms-data/files/carrier-rif>
<https://www.resdac.org/cms-data/files/carrier-lds>

Physician Compare Datasets

Overview: The Centers for Medicare & Medicaid Services (CMS) provides official datasets for the Medicare.gov [Physician Compare website](#) to give you useful information about groups, individual physicians, and other clinicians currently enrolled in Medicare.

- **Physician Compare National Downloadable File:** This file contains general information about individual eligible professionals (EPs) such as demographic information and Medicare quality program participation.
- **Physician Compare 2016 Individual EP Public Reporting:** This file contains Physician Quality Reporting System (PQRS) and non-PQRS Qualified Clinical Data Registry (QCDR) measure performance rates reported by clinicians.
- **Physician Compare 2016 Group Public Reporting:** This file contains Physician Quality Reporting System (PQRS) and non-PQRS Qualified Clinical Data Registry (QCDR) measure performance rates reported by groups.
- **Physician Compare 2016 Group Public Reporting - Patient Experience:** This file contains the Consumer Assessment of Healthcare Providers and Systems (CAHPS) for PQRS measure performance rates reported by groups.
- **Physician Compare Clinician Utilization Data:** This is the first set utilization data for clinicians publicly reported through Physician Compare. This is a subset of the 2015 Healthcare Common Procedure Coding System (HCPCS) codes.

Access: Publicly downloadable

Because of data use agreements with data vendors, not all data on Physician Compare can be shared in this downloadable file. For questions about the Downloadable Database or public reporting, contact PhysicianCompare@Westat.com.

Additional detailed information

<https://data.medicare.gov/data/physician-compare>

Physician Shared Patient Patterns Data

Overview: CMS offers Physician Shared Patient Patterns Data which is briefly [described](#) as useful for looking at providers who share relationships with common patients. Sharing is defined as when "An organization or provider participating in the delivery of health services to the same patient within a 30 days, 60 days, 90 days and finally a 180 day period after another organization or provider participated in providing health services to the same patient."

Access: Publicly downloadable

Available: 2009-2015

Additional detailed information

<https://www.cms.gov/Regulations-and-Guidance/Legislation/FOIA/Referral-Data-FAQs.html>

Provider of Services (POS) Current Files

Overview: The POS file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information. The data are collected through the Centers for Medicare & Medicaid Services (CMS) Regional Offices. The file contains an individual record for each Medicare-approved provider and is updated quarterly. The data is an invaluable resource to a variety of stakeholders, including researchers and application developers.

This file has previously only been available to the public for a fee. Now, the Provider of Services (POS) files are being made available as free downloads. The most current four quarters will be posted to this website and the December quarter for each previous year will be available on the left margin. This file includes:

- provider number
- provider demographics
- facility size
- facility staffing

Access: Publicly downloadable at <http://www.nber.org/data/provider-of-services.html>

Available: 1984 – 2017

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html>

Provider Specific Data for Public Use

Overview: The Fiscal Intermediary maintains the Provider Specific File (PSF). The file contains information about the facts specific to the provider that affects computations for the Prospective Payment System. The Provider Specific files in text format are located in the Download section below for the following provider-types:

- Inpatient
- Skilled Nursing Facility
- Home Health Agency
- Hospice
- Inpatient Rehab
- Long Term Care
- Inpatient Psychiatric Facility
- Outpatient

Access: Publicly downloadable

Additional detailed information

https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ProspMedicareFeeSvcPmtGen/psf_SAS.html

Skilled Nursing Facility (SNF) MEDPAR Limited Data Set (LDS)

Overview: MedPAR consolidates Inpatient Hospital or Skilled Nursing Facility (SNF) claims data from the National Claims History (NCH) files into stay level records. The accumulation of claims submitted for the period commencing on a beneficiary's date of admission to an inpatient hospital or SNF and ending on the beneficiary's date of discharge from that hospital or SNF represents one stay. In the case of a SNF stay where the beneficiary has not yet been discharged and remains a patient, the claims submitted between the admission date to the SNF through the time of the MedPAR file creation, represent one stay. A stay record may represent one or more final action claims.

NOTE: Any given MedPAR file represents a static snapshot of a specific stay at the time the data was sourced from NCH. As such, any given stay record on a specific update of MedPAR DOES NOT NECESSARILY represent the final coding and/or payment information for that stay because if subsequent adjustments to the claims that comprise the stay occur after MedPAR is run, they will not be reflected on the file.

Access: Please see *Accessing CMS Limited Data Set Files*

Availability: FY 2002 – FY 2017

Note: Regarding the 1999 File - The file contains corrupted data in data element "Admission of Date of Death Interval" positions 528-532. Corruption occurred during the Y2K conversion of legacy data files. At present, there are no plans to correct the data.

Cost: \$650 per year

Media: CD ROM

Data Format: Comma separated variable block (CSV) with SAS® read-in program

Size: 1,063 Mb. NOTE: File may be too large to open on a desktop.

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/SkilledNursingFacilityMEDPARLDS.html>

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 detailed information

<https://www.resdac.org/cms-data/file-availability>

Standard Analytical Files – LDS

Overview: LDS Standard Analytic Files (SAFs), also known as Medicare claims files, are available for each claim type (Inpatient, Outpatient, Skilled Nursing Facility, Home Health Agency, Hospice, Carrier and Durable Medical Equipment).

See below for the Data Dictionaries:

Current Data Dictionaries	Version	Format
Carrier Data Dictionary	K	Excel
Durable Medical Equipment Data Dictionary	K	Excel
Home Health Data Dictionary	K	Excel
Hospice Data Dictionary	K	Excel
Inpatient Data Dictionary	K	Excel
Outpatient Data Dictionary	K	Excel
Skilled Nursing Facility Data Dictionary	K	Excel

Data Dictionaries 2012 - 2015	Version	Format
Durable Medical Equipment Data Dictionary	J	PDF
Home Health Data Dictionary	J	PDF
Hospice Data Dictionary	J	PDF
Inpatient Data Dictionary	J	PDF
Outpatient Data Dictionary	J	PDF
Physician Supplier Part B Data Dictionary	J	PDF
Skilled Nursing Facility Data Dictionary	J	PDF

Historical Data Dictionaries	Version	Format
Durable Medical Equipment Data Dictionary	Pre-2011	PDF

Home Health Data Dictionary	Pre-2011	PDF
Hospice Data Dictionary	Pre-2011	PDF
Inpatient Data Dictionary	Pre-2011	PDF
Outpatient Data Dictionary	Pre-2011	PDF
Physician Supplier Part B Data Dictionary	Pre-2011	PDF
Skilled Nursing Facility Data Dictionary	Pre-2011	PDF

Access: Please see *Accessing CMS Limited Data Set Files*

Additional detailed information

<https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/StandardAnalyticalFiles.html>

Administrative Provider Data

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:

<https://hcupnet.ahrq.gov/#setup>

All data available on CD-ROM

For additional information:

Agency for Healthcare Research and Quality
Office of Communications and Knowledge Transfer
5600 Fishers Lane, 7th Floor
Rockville, MD 20857

The public inquiries number is: (301) 427-1104.

Email: hcup@ahrq.gov

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
NEDS	Annually from 2006
NIS	Annually from 1988
KID	Every three year from 1997
SASD	Annually from 1997
SEDD	in pilot phase and currently unavailable

- 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.
 - NEDS** national estimates emergency department visits across the country
 - 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

Additional detailed information

<https://www.hcup-us.ahrq.gov/>

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:

https://www.cdc.gov/nchs/nhds/nhds_questionnaires.htm

NHDS Data on CD-ROM: Contact the Hospital Care Team at nhds@cdc.gov to get copies of the 1970-1978 and/or 1979-2007 multiyear CD-ROMs.

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 2010

Population: From 1988-2007 the NHDS collected data from a sample of approximately 270,000 inpatient records acquired from a national sample of about 500 hospitals. From 2008 to 2010 the sample size was reduced to 239. Only hospitals with an average length of stay of fewer than 30 days for all patients, general hospitals, or children's general hospitals are included in the survey. Federal, military, and Department of Veterans Affairs hospitals, as well as hospital units of institutions (such as prison hospitals), and hospitals with fewer than six beds staffed for patient use, are excluded.

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)

Population Survey Data

Area Health Resource File (AHRF)

Overview: AHRF 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 AHRF System. It is a database containing more than 7,000 variables for each of the nation's counties. AHRF contains information on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. The AHRF System is a synthesis of data from over 50 different primary source files. Quality Resource Systems maintains the AHRF 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:
<https://datawarehouse.hrsa.gov/data/dataDownload.aspx>
Full data may be obtained on CD-ROM, magnetic tape, and 3480 Cartridges.
- 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:** Can be linked to other datasets via geographic identifiers

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:
https://www.cdc.gov/brfss/annual_data/annual_data.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

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: Need application

Researchers may for studies that involve multiple states by submitting a proposal to CDC. The submission process is outlined below. Please refer to the [Proposal Guidelines\[PDF – 109KB\]](#) for more detailed instructions.

Proposal submissions must include:

[PRAMS Application Form\[PDF – 937 KB\]](#).

Project Abstract of 350 words or less.

Signed [Data Sharing Agreement\[PDF – 34KB\]](#) (including signatures of all authors).

Proposal reviews are conducted once a month. Complete proposals must be received by close of business on the 1st day of the month to be included in the review. Proposals received after the 1st will be reviewed the following month. The application form, abstract, and data sharing agreement should be combined into a single pdf file. File name should be the Principal Researcher's last name. Submissions should be sent to PRAMSProposals@cdc.gov. Incomplete submissions will be returned without review.

Researchers will receive an e-mail regarding the status of their proposal approximately 2 weeks after the review date. Disapproved proposals will be returned with comments. Approved proposals will be forwarded to the PRAMS states for review. Data sets for approved proposals are issued approximately 6–8 weeks following the review date.

Cost: No charge

Updated: 1988 – 2016

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:

https://thedataweb.rm.census.gov/ftp/cps_ftp.html

Additional contact information:

<https://www.census.gov/programs-surveys/cps/data-detail.html>

Contact dsd_ferrett@census.gov

Cost: CPS data are free and publicly available

Updated: Annually with monthly supplements since 1940

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

Employer Health Benefits Survey

Overview: The Kaiser Family Foundation and the Health Research and Educational Trust have conducted this annual survey since 1999. The archives of the Employer Health Benefits Survey include these surveys and a small business supplement of the 1998 survey conducted by the Foundation. The survey was previously conducted by KPMG Peat Marwick from 1991-1998 and the Health Insurance Association from 1987-1991. The survey, which is traditionally released each Fall, tracks trends in employer health insurance coverage, the cost of that coverage, and other topical health insurance issues. Findings are based on a nationally representative survey of public and private employers, including those who respond to the full survey and those who indicate only whether or not they provide health coverage. Firms surveyed range in size from three to more than 300,000 employees.

- Access:** Survey data are readily available in report form from the Kaiser Family Foundation Web site:
<https://www.kff.org/health-costs/report/employer-health-benefits-annual-survey-archives/>
- 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

Health and Medical Care Archive (HMCA)

Overview: The Community Tracking Study (CTS) is a large-scale longitudinal investigation of health system change and its effects on people. CTS is a project of the [Center for Studying Health System Change \(HSC\)](#). Designed to track a cohort of American communities at two-year intervals beginning in 1996, this major research effort, sponsored by The Robert Wood Johnson Foundation (RWJF), is gathering information to monitor and understand the evolution of health care in the United States. CTS is investigating the ways in which hospitals, health plans, physicians, safety net providers, and other provider groups are restructuring their systems, and the forces driving the organizational change. Additionally, the project is tracking health insurance coverage, access to care, use of health services, health care costs, and perceived quality of health care. Much of the information collected by CTS comes from nationally representative surveys of households (CTS Household Survey), health plans (RWJF Employer Health Insurance Survey), physicians (CTS Physician Survey), and Healthcare for Communities (National Survey of Alcohol, Drug, and Mental Health Problems) that are conducted by HSC.

Access: All HMCA data and documentation files are freely available to the general public. <https://www.kff.org/health-costs/report/employer-health-benefits-annual-survey-archives/>

To obtain restricted variables in their original form, analysts must agree to the terms and conditions of a restricted data use agreement. These legally binding agreements place strict controls on who may access restricted data and how the data are used, stored, and conveyed.

Contact information

Voice: (734) 615-7957

Fax: (734) 647-8700

apienta@umich.edu (Amy Pienta, HMCA Director)

aidudley@umich.edu (Amanda Dudley, HMCA manager)

hmca@icpsr.umich.edu (Both Amy Pienta and Amanda Dudley)

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

hrsquestions@umich.edu

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>

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

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 (LSOAI). Baseline data are collected with the SOA and SOAI, supplements to the National Health Interview Survey (NHIS). The LSOA is comprised of three follow-up interviews and the LSOAI 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 LSOAI data are available publicly through the NCHS Web-site:
<https://www.cdc.gov/nchs/lsa/lsa2.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: lsa@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+
- SOAI:** sample of 9,447 people age 70+
- LSOAI:** 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:
https://meps.ahrq.gov/data_stats/download_data_files.jsp

Additional information and research specific questions available:

MEPS Project Director
 Medical Expenditure Panel Survey
 Agency for Healthcare Research and Quality
 5600 Fishers Lane
 Rockville, MD 20857
mepsprojectdirector@ahrq.hhs.gov
 (301) 427-1406

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 \approx 11,000 households
NHC	sample consists of \approx 815 nursing homes and 6,000 residents
MPC	sample consists of \approx 22,000 providers to supplement HC information
IC	annual survey of \approx 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 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 via:
<https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>
Additional data for research purposes is available after submission of a research proposal.

For more information:

Division of Health Interview Statistics
National Center for Health Statistics
3311 Toledo Rd, Room 2217
Hyattsville, MD 20782-2064
(301) 458-4901
(301) 458-4001
nhis@cdc.gov

Cost: Free (Publicly Held) for download

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 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 <https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00216>
- 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: <https://simba.isr.umich.edu/data/data.aspx>

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

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:
<https://archive.ahrq.gov/data/safetynet/#Data>

For more information:

Agency for Healthcare Research and Quality
 540 Gaither Road
 Rockville, MD 20850
 Tel: 301-427-1364
 E-mail: <https://info.ahrq.gov>

Cost: Free (Publicly Available)

Updated: 2003

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:
https://www.cdc.gov/nchs/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 at the NBER website
<http://www.nber.org/data/survey-of-income-and-program-participation-sipp-data.html>

More information

Phone: 1.888.245.3076

Email: census.sipp@census.gov

Cost: Free

Updated: Panel and wave from 1984 to 2014

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

Survey Methods: SIPP 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

The Commonwealth Fund

Overview: The Commonwealth Fund and the Center for Studying Health System Change offers a window into the state of retiree health coverage in the U.S. Based on findings from a survey of employee benefit managers in public and private firms that offer health benefits to retired workers, this resource provides answers to such questions as:

- What percentage of U.S. companies offer retiree health benefits?
- How have private and public employers responded to enactment of Medicare's Part D prescription drug benefit?
- How does the size of plan deductibles vary among Medicare-age retirees?
- Do employers plan to continue offering the same level of retiree coverage in the future?
- What percentage of firms are planning to increase retired workers' share of insurance premiums?

Additional detailed information

<https://www.commonwealthfund.org/publications/surveys>

Survey Insurance Data

HealthLeaders – InterStudy

Overview: On both a national and a local level, HealthLeaders-InterStudy are the experts on health plans and healthcare markets. **Health Plan Data and Analysis** - Information on benefit designs, contracting strategies, financial performance, and enrollment information by state and MSA. By identifying key health plan trends that affect your business, Health Plan Data & Analysis allows you to create your most comprehensive strategic plan AND sales strategies at state and local levels. **Market Overviews** - Know the major healthcare players in the 75 most significant metro markets By giving you a detailed analysis of each local healthcare market, each Market Overview allows you to plan your specific strategy and have confidence that you know the situation, the players, and the trends. **Pharmacy Benefit Evaluator** - Managed care market information for the pharmaceutical industry With a comprehensive analysis of every managed care plan's pharmacy benefit, Pharmacy Benefit Evaluator gives pharmaceutical planning, marketing, and sales a comprehensive knowledge of the managed care market. **HMO Financial Analyzer** - Understand HMO financial and membership metrics. Supported by data on every publicly traded and private HMO in the country, HMO Financial Analyzer lets you understand HMO performance in detail. **Medicare Outlook** - Detailed Medicare market insight & data. By providing an in-depth look at the market developments within the Medicare landscape, Medicare Outlook is a necessary resource of data and analysis key to understanding this dynamic managed care segment. **Employer Vantage** - Comprehensive employer health benefit database. As the only source to reveal the relationships an employer has with its contracted insurance carriers and service providers, Employer Vantage provides organizations with the information to develop effective business strategies to employers of all sizes.

Additional detailed information

<https://decisionresourcesgroup.com/solutions/hli-us-market-access-dynamics/>

Survey Provider Data

American Hospital Association (AHA) Annual Survey

Overview: The AHA Annual Survey Database is a comprehensive census of United States hospitals based on the *AHA Annual Survey of Hospitals*. The Database is a reliable resource for health service research and trends analyses, and it offers a snapshot of hospital-specific data on approximately 6,500 hospitals and 400-plus systems, including as many as 1,000 data fields covering organizational structure, personnel, hospital facilities and services, and financial performance. Includes a documentation book with a Summary of Changes that identifies open, closed, and merged hospitals compared to the prior year. The database is released annually in October.

Access: Information is in ASCII and Microsoft Access format on CD-ROM Available through order from the AHA at:
<http://www.ahadata.com/aha-annual-survey-database-asdb>

For more information contact:

155 N. Wacker Dr.

Chicago, Illinois 60606

[312.422.3000](tel:312.422.3000)

800 10th Street, N.W.

Two CityCenter, Suite 400

Washington, DC 20001-4956

[202.638.1100](tel:202.638.1100)

Contact Health Forum: 866-375-3633 or ahadatainfo@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

Integrated Healthcare Delivery System (IHDS+) Database

Overview: The Dorenfest Complete Integrated HEALTHCARE DELIVERY SYSTEM PLUS (IHDS+) DATABASE™ is a market intelligence tool that has profiled Hospital and Integrated Delivery System IT efforts from 1986 up to two years before the current year. From 1986 to 1995, Sheldon I. Dorenfest & Associates, Ltd. tracked demographic and information systems data for all non-federal acute care facilities with 100 beds or more in THE DORENFEST 3000+ DATABASE™.

In 1998, THE FIRST DORENFEST COMPLETE IHDS+ DATABASE™, which included all integrated healthcare delivery systems (IHDSs) that own at least one short term, acute care, non-federal hospital with at least 100 beds, was published. This product contains 1,467 IHDSs and data on over 32,000 facilities. The database provides a comprehensive picture of healthcare delivery in the United States. This was followed by updated versions in 1999, 2000, 2001, 2002, and 2003 (respectively, the SECOND, THIRD, FOURTH, FIFTH, and SIXTH DORENFEST COMPLETE IHDS+ DATABASE™).

Information in the database includes:

- Market segmentation and size statistics
- IT purchase plans for healthcare organizations
- Software, hardware, and infrastructure installed throughout all facilities within each IDN

For additional general information contact:

Email: foundation@himss.org

Phone: 312-664-4467

Healthcare Integrity and Protection Data Bank

Overview: The Secretary of the U.S. Department of Health and Human Services, acting through the Office of Inspector General (OIG), was directed by the Health Insurance Portability and Accountability Act of 1996 to create the Healthcare Integrity and Protection Data Bank (HIPDB) to combat fraud and abuse in health insurance and health care delivery. Health care fraud burdens the nation with enormous financial costs and threatens the quality of health care and patient safety. Estimates of annual losses due to health care fraud range from 3 to 10 percent of all health care expenditures -- between \$30 billion and \$100 billion based on estimated 1997 expenditures of over \$1 trillion.

Access to information in the HIPDB is available to entities that meet the eligibility requirements defined in Section 1128E of the Social Security Act and the HIPDB regulations. In order to access information, eligible entities must first register with the Data Bank.

Additional detailed information

<https://www.npdb.hrsa.gov/resources/hipdbArchive.jsp>

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

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: <https://www.jointcommission.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:

https://www.jointcommission.org/about_us/data_mart.aspx

For additional general information contact:

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

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: <https://www.mgma.com/data>

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 surgery 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)

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**).

HOS PUFs are available at no cost and can be downloaded directly from <http://www.hosonline.org/en/data-dissemination/research-data-files/>

A signed Data Use Agreement with CMS is required to obtain either LDS or RIF data files. The 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](http://www.resdac.umn.edu) (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.

Available: 1998 – 2016

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®.

Medicare Physician Identification and Eligibility Registry (MPIER)

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. This file includes:

- physician UPIN number,
- name,
- date of birth and death,
- medical school and year of graduation,
- state license number,
- primary and secondary specialties,
- billing and business zipcodes,
- affiliations (solo practitioner or member of a group practice)

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*.

Available: CY 2003 – 2007 Q2

Additional Information: For access to variable descriptions see:

<https://www.resdac.org/cms-data/files/mpier>

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 via https://www.cdc.gov/nchs/ahcd/datasets_documentation_related.htm

For more information:

Ambulatory and Hospital Care Statistics Branch
National Center for Health Statistics
3311 Toledo Road
Hyattsville, Maryland 20782
301-458-4600

Cost: Free

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)

Consumer Assessment of Healthcare Providers and Systems (CAHPS) Database

Overview: The CAHPS Database (formerly known as the National CAHPS Benchmarking Database) is AHRQ's repository for data from selected CAHPS surveys. The primary purpose of the CAHPS Database is to facilitate comparisons of CAHPS survey results by and among survey users. To support the use of CAHPS survey results, the CAHPS Database offers products and services that include a Chartbook, an Online Reporting System, custom analyses, and data for research purposes.

The CAHPS Database currently includes:

- CAHPS Health Plan Survey Data—The CAHPS Health Plan Database contains results for the Commercial and Medicaid Survey submitted by various sponsors, including public and private employers, State Medicaid agencies, State Children's Health Insurance Programs (CHIP), and individual health plans. The data sets date back to 1998.
- CAHPS Clinician & Group Survey Data—The CAHPS Clinician & Group Database contains results for multiple versions of this survey submitted by various sponsors, including practice sites, medical groups, and regional health collaboratives. Data on patient experience with doctors and medical groups can be viewed by region, physician specialty, practice ownership and affiliation, number of visits by practice, and survey mode. The data sets date back to 2010.

De-Identified Data Requests

Researchers interested in obtaining de-identified CAHPS survey data from the CAHPS Database must submit a completed Research Abstract Form and a signed Data Release Agreement.

- Information and Instructions for [Requesting De-identified Data](#) (PDF, 128 KB)—Provides information and application instructions for researchers interested in obtaining de-identified data files from the CAHPS Database.
- [Research Abstract Form](#) (Word Version, 129 KB)—Researchers requesting data must submit a completed form outlining the proposed use of the data.
- For questions or to apply for de-identified CAHPS Data:
Email CAHPSDatabase@westat.com (Subject Line: De-identified CAHPS data).

Requests will be fulfilled year-round. On average, the time from receipt of request to release of data is approximately 45 days

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.

Additional detailed information

<https://www.ahrq.gov/cahps/cahps-database/data-research/index.html>

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: NHHCS data releases are available for public use via the NHIS web site:
<https://www.cdc.gov/nchs/nhhcs/index.htm>

For more information:

Long-term Care Statistics Branch at (301) 458-4747

Cost: Free

Updated: Periodically from 1992 to 2000, and 2007

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 via the NHIS web site:
https://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm

For more information:

Ambulatory and Hospital Care Statistics Branch
National Center for Health Statistics
3311 Toledo Road
Hyattsville, Maryland 20782
301-458-4600

Cost: Free

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 Nursing Home Survey (NNHS)

Overview: NNHS 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:** NNHS data releases are available for public use via the NHIS web site:
https://www.cdc.gov/nchs/nnhs/nnhs_questionnaires.htm
- Cost:** Free
- Updated:** Periodically from 1973 to 1999, and 2004
- 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 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 from the NPDB website:
<https://www.npdb.hrsa.gov/resources/publicData.jsp>

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

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 via the NHIS web site:
<https://www.cdc.gov/nchs/nsas/index.htm>

Cost: Free

Updated: 1994-1996, and 2006

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)

Other Data

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 information : <https://aishealth.com/category/datapoint/>

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>

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: HSCdataHelp@hschange.org.

Cost: Public use and restricted data files are free

Updated:	Household Survey	1996-97, 1998-99, 2000-01, 2003-04, 2007, 2010
	Followback Survey	1996-97, 1998-99
	Physician Survey	1996-97, 1998-99, 2000-01, 2004-05, 2008
	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

HcPro (formerly National Health Information NHI)

Overview: HCPro's Healthcare Marketplace is dedicated to bringing you the highest quality healthcare management resources available. Our products and services keep you informed of the latest regulations and standards, provide insightful examples and advice from industry experts, and offer practical and effective solutions so that you can do your job better. HCPro is a pioneer in the delivery of Web-based information, resources, and content on "The Business of Healthcare" to managers in the healthcare industry. HCPro developed a series of unique Internet-based solutions including commercial Web sites, online resource centers, electronic product distribution, electronic site licenses, and strategic distribution alliances. NHI, of Atlanta, GA, is a business-to-business healthcare publisher with subscription newsletters and books in managed care, and also in several of HCPro's core markets, such as health information management and long-term care. HCPro is purchasing all the assets of NHI, including 13 newsletters and ancillary products, several manuals and one directory.

- **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:

Tel: 800/650-6787

e-mail customerservice@hcpro.com

<http://www.hcpro.com/services/nhi/>

Health Resources and Services Administration – Geospatial Data Warehouse

Overview: The Health Resources and Services Administration (HRSA), an agency of the U.S. Department of Health and Human Services, is the primary Federal agency for improving access to health care services for people who are uninsured, isolated or medically vulnerable. Comprising six bureaus and 12 offices, HRSA provides leadership and financial support to health care providers in every state and U.S. territory. HRSA grantees provide health care to uninsured people, people living with HIV/AIDS, and pregnant women, mothers and children. They train health professionals and improve systems of care in rural communities. HRSA oversees organ, tissue and blood cell (bone marrow and cord blood) donation and vaccine injury compensation programs, and maintains databases that protect against health care malpractice and health care waste, fraud and abuse.

Data sources at

<https://datawarehouse.hrsa.gov/data/aboutData/dataSources.aspx>

National Association of Health Underwriters

Overview: There were as many as 45 million Americans that went without health insurance coverage at some point during 2006. Many of these people may not have obtained needed medical care as a result, and all took a significant financial gamble. The National Association of Health Underwriters, a professional association of more than 20,000 health insurance agents, brokers and benefit specialists, is extremely concerned about the problem of the uninsured. One of our primary goals is to help ensure that Americans have access to appropriate health coverage. To help address this problem, we have developed the Health Care Coverage Options Database (Database), an online tool to help make American health care consumers aware of all of the coverage options available to them. The Database contains information about private health insurance coverage, as well as the many public and private programs available to Americans to help them obtain the medical care they need. Policymakers, the media and other interested parties can also use the Database as a means of comparing health care coverage options for Americans on a state-by-state basis. The Database is broken up in to four components:

Employer-Based Health Insurance Coverage The majority of Americans have group health insurance coverage through either their employer, or the employer of a family member. This section is designed for consumers who need more information about what requirements apply to group health insurance coverage or need to know what their options are if they are about to lose access to their group coverage. It also provides contact information for the state and federal regulators who oversee private group health insurance in America, in case consumers have questions or need to report a problem. Consumers who are looking to purchase a group health insurance product can go to NAHU's Find An Agent feature to find an NAHU member in their area to assist them in obtaining coverage.

Individual Health Insurance Coverage Many Americans who do not have access to group health insurance coverage, like self-employed people, early retirees,

Additional detailed information

<https://nahu.org/>

National Cancer Institute – Surveillance Epidemiology and End Result (SEER)

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.

There are two ways to obtain the research data: both require access through an Internet connection. Please read through each option below to decide which method best suits your needs. A signed [SEER Research Data Agreement](#) is required in both cases.

For additional SEER information including a complete listing of data available go to: <https://seer.cancer.gov/data/>

For additional general information contact:

If you have been unable to find the information you need in our [online resources](#), you can [contact NCI's SEER Program](#). Answers to specific statistical questions may involve complex databases. Please allow up to 2 weeks for a response.

[Cancer Stat Facts](#), a collection of statistical summaries for a number of common cancer types, provide a quick overview of frequently-requested cancer statistics.

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/resources/reports>

For more information, file [online request form](#)

Quality Compass National Committee for Quality Assurance (NCQA)

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/hedis-quality-measurement/quality-measurement-products/quality-compass>
Data are available either SAS or Excel

For additional information:

NCQA
1100 13th St., NW, Third Floor
Washington, D.C. 20005
Customer Support Tel: 888-275-7585
Tel: 202-955-3500
Fax: 202-955-3599
E-mail: Customersupport@ncqa.org

Cost: Please see [here](#) for detailed information

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

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. As of October 16, 2006, **Solucient**, LLC was acquired by Thomson Corp. 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>

United States Department of Labor – Bureau of Labor Statistics (BLS)

Overview: The Bureau of Labor Statistics annually reports on the number of workplace injuries, illnesses, and fatalities. Such information is useful in identifying industries with high rates or large numbers of injuries, illnesses and fatalities both nationwide and separately for those States participating in this program. Since 1972, the survey has reported annually on the number of workplace injuries and illnesses in private industry and the frequency of those incidents. With the 1992 survey, BLS began collecting additional information on the more seriously injured or ill workers in the form of worker and case characteristics. At that time, BLS also initiated a separate Census of Fatal Occupational Injuries to count these tragic events more effectively than had been possible in the survey.

The BLS safety and health statistical system, therefore, presents three distinct types of data.

- Summary data – This reports the number and rate of injuries and illnesses by industry.
- Case and demographic data – which provides additional details on the worker injured, the nature of the disabling condition, and the event and source producing that condition for those cases that involve one or more days away from work.
- Fatality data – which provides information on 28 separate data elements including information on the worker, the fatal accident, and the machinery or equipment involved.

Additional information is also available about the history of the BLS safety and health statistical programs.

The National Compensation Survey (NCS) provides comprehensive measures of occupational earnings; compensation cost trends, benefit incidence, and detailed plan provisions. Detailed occupational earnings are available for metropolitan and non-metropolitan areas, broad geographic regions, and on a national basis. The index component of the NCS (ECI) measures changes in labor costs. Average hourly employer cost for employee compensation is presented in the ECEC. Related BLS programs

- 1) Consumer Price Indexes
- 2) National Compensation Survey – Compensation Cost Trends
- 3) Labor Force Statistics from the Current Population Survey
- 4) National Compensation Survey – Benefits

Access and Further Information:

<https://www.bls.gov/data/>

World Health Organization (WHO)

Overview: World health statistics 2007 presents the most recent health statistics for WHO's 193 Member States. This third edition includes a section with 10 highlights of global health statistics for the past year as well as an expanded set of 50 health statistics. World Health Statistics 2007 has been collated from publications and databases produced by WHO's technical programmes and regional offices. The core set of indicators was selected on the basis of their relevance to global health, the availability and quality of the data, and the accuracy and comparability of estimates. The statistics for the indicators are derived from an interactive process of data collection, compilation, quality assessment and estimation occurring among WHO's technical programmes and its Member States. During this process, WHO strives to maximize the accessibility, accuracy, comparability and transparency of health statistics.

DATABASES The WHOSIS database: includes data from World Health Statistics and the World Health Report These are the latest Core Health Indicators from WHO sources, including 'World Health Statistics' and 'The World Health Report'. With this query system, you can construct tables for any combination of countries, indicators and years. Causes of Death database Causes of death by country, year, sex, cause and age group. Number of registered deaths, infant deaths, estimated completeness and coverage of mortality data for latest year. WHO Global InfoBase Online The WHO Global InfoBase Online is a data warehouse that collects stores and displays information on chronic diseases.

Statistical information can be downloaded or hard copies can be ordered for a nominal charge.

Access and Further Information:

<http://www.who.int/whosis/en/>