

# **Compendium of Federal Data Sources to Support Health Workforce Analysis, August 2017**



Health Resources and Services Administration  
Bureau of Health Workforce  
National Center for Health Workforce Analysis

The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS), provides national leadership in the development, distribution, and retention of a diverse, culturally competent health workforce that can adapt to the population's changing health care needs and provide the highest quality care for all. The agency administers a wide range of training grants, scholarships, loans, and loan repayment programs that serve as a catalyst to advance changes in health professions training that respond to the evolving needs of the health care system.

The National Center for Health Workforce Analysis (National Center) informs public and private-sector decision-making related to the health workforce by expanding and improving health workforce data, disseminating workforce data to the public, improving and updating projections of the supply and demand for health workers, and conducting analyses of issues important to the health workforce.

For more information about the National Center, email us at [healthworkforcecenter@hrsa.gov](mailto:healthworkforcecenter@hrsa.gov) or visit our website at <https://bhw.hrsa.gov/health-workforce-analysis>.

## Table of Contents

Introduction .....	5
Table 1: Summary List of Federal Data Sources for Health Workforce Analysis .....	7
<b>Agency for Healthcare Research and Quality (AHRQ)</b>	
Healthcare Cost and Utilization Project (HCUP).....	12
Consumer Assessment of Healthcare Providers and Systems (CAHPS) .....	14
Medical Expenditure Panel Survey (MEPS) .....	16
<b>Bureau of Labor Statistics (BLS)</b>	
American Time Use Survey (ATUS).....	19
Current Employment Statistics (CES) Survey .....	21
Employment Projections (EP) Program .....	22
Occupational Employment Statistics (OES) Survey .....	24
<b>Census Bureau</b>	
American Community Survey (ACS) .....	26
Current Population Survey (CPS) .....	29
Economic Census.....	31
Survey of Business Owners and Self-Employed Persons (SBO) .....	33
Survey of Income and Program Participation (SIPP).....	35
<b>Centers for Disease Control and Prevention</b>	
Behavioral Risk Factor Surveillance System (BRFSS).....	37
National Ambulatory Medical Care Survey (NAMCS) .....	39
National Health and Nutrition Examination Survey (NHANES) .....	41
National Hospital Ambulatory Medical Care Survey (NHAMCS).....	43
National Hospital Discharge Survey (NHDS).....	45
National Study of Long-Term Care Providers (NSLTCP).....	47
National Survey of Residential Care Facilities (NSRCF) .....	49
National Youth Tobacco Survey (NYTS) .....	51
Pregnancy Risk Assessment Monitoring System (PRAMS).....	53
Youth Risk Behavior Surveillance System (YRBSS) .....	55
<b>Centers for Medicare and Medicaid Services (CMS)</b>	
Health Outcomes Survey (HOS) .....	57
Medicare Claims Data .....	59
Medicaid Claims Data (national Medicaid Analytic eXtract [MAX] Data) .....	62

Medicare Current Beneficiary Survey (MCBS) .....	65
National Provider Identifier (NPI) File.....	67
<b>Data Resource Center for Child and Adolescent Health (DRC)</b>	
National Survey of Children’s Health (NSCH) .....	69
National Survey of Children with Special Health Care Needs (NS-CSHCN) ..	71
<b>Health Resources and Services Administration</b>	
HRSA Data Warehouse .....	72
Area Health Resources Files (AHRF).....	73
Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps) .....	76
National Sample Survey of Registered Nurses (NSSRN).....	78
National Sample Survey of Nurse Practitioners (NSSNP) .....	80
<b>National Center for Education Statistics (NCES)</b>	
Integrated Postsecondary Education Data System (IPEDS) .....	82
<b>Substance Abuse and Mental Health Services Administration (SAMHSA)</b>	
National Mental Health Services Survey (N-MHSS) .....	84
National Survey of Substance Abuse Treatment Services (N-SSATS) .....	86
National Survey on Drug Use and Health (NSDUH) .....	87
Appendix.....	88

## Introduction

The United States Government supports a wide range of surveys that collect data and information to inform public and private-sector decision-making. The Health Resources and Services Administration's National Center for Health Workforce Analysis (National Center) has reviewed existing surveys and sources of data and assessed the information that each can provide on the supply, utilization, access, need, and/or demand for health workers.

This compendium provides a summary of 38 Federal data sources that, while not established specifically to collect or present health workforce data, can be used to support health workforce analysis. The compendium is designed to help those interested in the health workforce to understand how these sources could be used for health workforce analysis. For each data source, this compendium provides the following:

- Lead federal agency
- Website
- Description of data source
- Relevance for health workforce analysis
- Geographical detail
- Availability

There are a number of other, non-Federal resources available (most of which are for purchase) that could be used for workforce analysis (e.g., the AMA Master File); although these are not summarized here since this report focuses only on Federal data sources. Additionally, the [NIH Health Services Research Information Central](#) website is a repository of several Federal and non-Federal databases that may be of use to health workforce researchers.

The National Center was established to promote improved and expanded health workforce data and analysis to inform public policies and private decision-making related to the health workforce. The National Center will periodically update this compendium and welcomes suggestions for additions to this list, as well as insights on the use of these data sources for health workforce analysis and studies.

For more information about the National Center, please visit the Center's website at <https://bhw.hrsa.gov/health-workforce-analysis> or email the Center at [healthworkforcecenter@hrsa.gov](mailto:healthworkforcecenter@hrsa.gov).

**Table 1: Summary List of Federal Data Sources for Health Workforce Analysis**

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
<b>Healthcare Cost and Utilization Project (HCUP)</b>	AHRQ, DHHS	<a href="https://www.ahrq.gov/research/data/hcup">https://www.ahrq.gov/research/data/hcup</a>	Health claims data	Health care encounter	Health care demand
<b>Consumer Assessment of Healthcare Providers and Systems (CAHPS)</b>	AHRQ, DHHS	<a href="https://www.ahrq.gov/cahps">https://www.ahrq.gov/cahps</a>	Patient survey	Individual	Health care quality, satisfaction, and access
<b>Medical Expenditure Panel Survey (MEPS)</b>	AHRQ, DHHS	<a href="http://meps.ahrq.gov/mepsweb">http://meps.ahrq.gov/mepsweb</a>	Household survey with component on sampled individual's health utilization data	Individual	Health care demand
<b>American Time Use Survey (ATUS)</b>	BLS, DOL	<a href="https://www.bls.gov/tus">https://www.bls.gov/tus</a>	Household survey	Individual	Provider supply; health care demand
<b>Current Employment Statistics (CES) Survey</b>	BLS, DOL	<a href="https://www.bls.gov/ces">https://www.bls.gov/ces</a> ; <a href="https://www.bls.gov/sae">https://www.bls.gov/sae</a>	Payroll/establishment survey (businesses and government agencies)	Total employment by industry	Total employment in health care establishments
<b>Employment Projections (EP) Program</b>	BLS, DOL	<a href="https://www.bls.gov/emp">https://www.bls.gov/emp</a>	Projection model	Occupation	Employment projections for health occupations
<b>Occupational Employment Statistics (OES) Survey</b>	BLS, DOL	<a href="https://www.bls.gov/oes">https://www.bls.gov/oes</a>	Establishment mail survey	Occupation	Provider supply
<b>American Community Survey (ACS)</b>	U.S. Census Bureau, U.S. Department of Commerce	<a href="https://www.census.gov/programs-surveys/acs">https://www.census.gov/programs-surveys/acs</a>	Household survey	Household and individual	Provider supply
<b>Current Population Survey (CPS)</b>	U.S. Census Bureau, U.S. Department of Commerce; BLS, DOL	<a href="https://www.census.gov/programs-surveys/cps.html">https://www.census.gov/programs-surveys/cps.html</a>	Household survey	Individual	Provider supply
<b>Economic Census</b>	U.S. Census Bureau, U.S. Department of Commerce	<a href="https://www.census.gov/programs-surveys/economic-census.html">https://www.census.gov/programs-surveys/economic-census.html</a>	Establishment survey	Establishment	Establishment characteristics; provider staffing

**Table 1: Summary List of Federal Data Sources for Health Workforce Analysis (cont'd.)**

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
<b>Survey of Business Owners and Self-Employed Persons (SBO)</b>	U.S. Census Bureau, U.S. Department of Commerce	<a href="https://www.census.gov/programs-surveys/sbo.html">https://www.census.gov/programs-surveys/sbo.html</a>	Company survey	Company	Company characteristics; provider staffing
<b>Survey of Income and Program Participation (SIPP)</b>	U.S. Census Bureau, U.S. Department of Commerce	<a href="https://www.census.gov/sipp">https://www.census.gov/sipp</a>	Household survey	Household and individual	Health care demand
<b>Behavioral Risk Factor Surveillance System (BRFSS)</b>	Office of Surveillance, Epidemiology, and Laboratory Services, CDC, DHHS	<a href="https://www.cdc.gov/BRFSS">https://www.cdc.gov/BRFSS</a>	Telephone survey	Individual	Health care demand and access
<b>National Ambulatory Medical Care Survey (NAMCS)</b>	National Center for Health Statistics, CDC, DHHS	<a href="https://www.cdc.gov/nchs/ahcd">https://www.cdc.gov/nchs/ahcd</a>	Office-based physician survey	Health care encounter and physician	Provider supply; health care demand
<b>National Health and Nutrition Examination Survey (NHANES)</b>	CDC, DHHS	<a href="https://www.cdc.gov/nchs/nhanes">https://www.cdc.gov/nchs/nhanes</a>	Survey	Individual	Health care demand
<b>National Hospital Ambulatory Medical Care Survey (NHAMCS)</b>	National Center for Health Statistics, CDC, DHHS	<a href="https://www.cdc.gov/nchs/ahcd">https://www.cdc.gov/nchs/ahcd</a>	Establishment survey (hospital outpatient and emergency departments)	Health care encounter	Health care demand
<b>National Hospital Discharge Survey (NHDS)</b>	National Center for Health Statistics, CDC, DHHS	<a href="https://www.cdc.gov/nchs/nhds">https://www.cdc.gov/nchs/nhds</a>	Establishment survey (hospitals)	Health care encounter	Health care demand
<b>National Study of Long-Term Care Providers (NSLTCP)</b>	CDC, DHHS	<a href="https://www.cdc.gov/nchs/nsltcp">https://www.cdc.gov/nchs/nsltcp</a>	Establishment survey	Establishment	Provider staffing; health care demand
<b>National Survey of Residential Care Facilities (NSRCF)</b>	National Center for Health Statistics, CDC, DHHS	<a href="https://www.cdc.gov/nchs/nsrcf">https://www.cdc.gov/nchs/nsrcf</a>	Survey	Individual health provider	Provider staffing; health care demand up to 2010 (questions now folded into NSLTCP)



**Table 1: Summary List of Federal Data Sources for Health Workforce Analysis (cont'd.)**

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
<b>National Youth Tobacco Survey (NYTS)</b>	CDC, DHHS	<a href="https://www.cdc.gov/tobacco/data_statistics/surveys/nyts">https://www.cdc.gov/tobacco/data_statistics/surveys/nyts</a>	Youth survey	Individual	Health care demand
<b>Pregnancy Risk Assessment Monitoring System (PRAMS)</b>	CDC, DHHS	<a href="https://www.cdc.gov/prams">https://www.cdc.gov/prams</a>	Survey of women after childbirth	Individual	Health care demand
<b>Youth Risk Behavior Surveillance System (YRBSS)</b>	CDC, DHHS	<a href="https://www.cdc.gov/healthyyouth/data/yrbss">https://www.cdc.gov/healthyyouth/data/yrbss</a>	Youth survey	Individual	Health care demand
<b>Health Outcomes Survey (HOS)</b>	CMS, DHHS	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS</a>	Medicare beneficiary survey	Individual	Health care demand
<b>Medicare Claims Data</b>	CMS, U.S. DHHS	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html</a>	Health claims data	Health care encounter	Provider supply; health care demand; health care access
<b>Medicaid Claims Data (Medicaid Analytic eXtract [MAX] Data)</b>	CMS, DHHS	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html">https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html</a>	Medicaid claims data	Individual	Health care demand
<b>Medicare Current Beneficiary Survey (MCBS)</b>	CMS, DHHS	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS">https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS</a>	Individual survey supplemented with claims data	Individual	Health care demand; health care access
<b>National Provider Identifier (NPI) File</b>	CMS, DHHS	<a href="https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand">https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand</a>	Enrollment in program by health care providers	Individual health provider	Provider supply

**Table 1: Summary List of Federal Data Sources for Health Workforce Analysis (cont'd.)**

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
<b>National Survey of Children's Health (NSCH)</b>	Data Resource Center for Child and Adolescent Health	<a href="http://www.childhealthdata.org/learn/NSCH">http://www.childhealthdata.org/learn/NSCH</a>	Household survey	Individual children	Health care demand
<b>National Survey of Children with Special Health Care Needs (NS-CSHCN)</b>	Data Resource Center for Child and Adolescent Health	<a href="http://www.childhealthdata.org/learn/NS-CSHCN">http://www.childhealthdata.org/learn/NS-CSHCN</a>	Household survey	Individual children	Health care demand
<b>Area Health Resource File (AHRF)</b>	National Center for Health Workforce Analysis, HRSA, DHHS	<a href="http://arf.hrsa.gov">http://arf.hrsa.gov</a>	Compilation of multiple data sources to populate information on U.S. counties	County	Provider supply
<b>Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps)</b>	HRSA, DHHS	<a href="https://bhw.hrsa.gov/shortage-designation">https://bhw.hrsa.gov/shortage-designation</a>	Health workforce shortage designations	Geographic areas; special populations	Health care demand
<b>HRSA Data Warehouse (HDW)</b>	HRSA, DHHS	<a href="https://datawarehouse.hrsa.gov/default.aspx">https://datawarehouse.hrsa.gov/default.aspx</a>	Compilation of multiple data sources to populate	Geographic areas; special populations; Establishment	Health care demand; Provider supply; Establishment characteristics; patient population
<b>National Sample Survey of Registered Nurses (NSSRN)</b>	HRSA, DHHS	<a href="https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSRN.aspx">https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSRN.aspx</a>	Survey	Individual registered nurse	Provider supply
<b>National Sample Survey of Nurse Practitioners (NSSNP)</b>	HRSA, DHHS	<a href="https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSNP2012.aspx">https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSNP2012.aspx</a>	Survey	Individual nurse practitioner	Provider supply
<b>Integrated Postsecondary Education Data System (IPEDS)</b>	National Center for Education Statistics, ED	<a href="https://nces.ed.gov/ipeds">https://nces.ed.gov/ipeds</a>	Establishment survey	Establishment	Provider supply
<b>National Mental Health Services Survey (N-MHSS)</b>	SAMHSA, DHHS	<a href="https://www.samhsa.gov/data/mental-health-facilities-data-nmhss">https://www.samhsa.gov/data/mental-health-facilities-data-nmhss</a>	Establishment survey	Establishment	Establishment characteristics; patient population

**Table 1: Summary List of Federal Data Sources for Health Workforce Analysis (cont'd.)**

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
<b>National Survey of Substance Abuse Treatment Services (N-SSATS)</b>	SAMHSA, DHHS	<a href="https://www.samhsa.gov/data/substance-abuse-facilities-data-nssats">https://www.samhsa.gov/data/substance-abuse-facilities-data-nssats</a>	Establishment survey	Establishment	Establishment characteristics; health care demand
<b>National Survey on Drug Use and Health (NSDUH)</b>	SAMHSA, DHHS	<a href="https://www.samhsa.gov/data/population-data-nsduh">https://www.samhsa.gov/data/population-data-nsduh</a>	Survey	Individual	Health care demand

Note: Federal Agency Acronyms

AHRQ - Agency for Healthcare Research and Quality

BLS - Bureau of Labor Statistics

CDC - Centers for Disease Control and Prevention

CMS - Centers for Medicare and Medicaid Services

DHHS - U.S. Department of Health and Human Services

DOL - U.S. Department of Labor

ED - U.S. Department of Education

HRSA - Health Resources and Services Administration

SAMHSA - Substance Abuse and Mental Health Services Administration

## Healthcare Cost and Utilization Project (HCUP)

**FEDERAL AGENCY:** Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.ahrq.gov/research/data/hcup>

**DESCRIPTION:** The Healthcare Cost and Utilization Project (HCUP) is a family of health care databases and related software tools and products. National databases include the Nationwide Inpatient Sample (NIS), the Kids' Inpatient Database (KID), and the Nationwide Emergency Department Sample (NEDS). The NIS, started in 1988, is the largest all-payer inpatient care database in the U.S., with data from approximately 8 million hospital stays from roughly 1,000 hospitals. The KID, started in 1997, is a nationwide sample of pediatric inpatient discharges. Started in 2006, the NEDS contains more than 25 million records for emergency department visits at about 1,000 U.S. community hospitals.

HCUP databases contain a core set of clinical and nonclinical data found in a typical discharge abstract, including patient demographics, listed diagnoses and procedures, discharge status, and charges for all patients, regardless of insurance status (insured or uninsured) and payer (e.g., Medicare, Medicaid, or private insurance).

State databases covering inpatient stays, ambulatory surgery, and emergency department use for participating states are also part of HCUP.

The sample size in national data files varies according to the type of health care captured by the data. State-specific data files, where available, generally include records for the universe of the event captured in the data file (e.g., all inpatient discharges in the state during the reporting period).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The HCUP focuses primarily on the use and cost of hospital services, specifically inpatient and emergency department services. The data in the HCUP are applicable to analyses that estimate or predict how demand for hospital inpatient and emergency care—and by extension, demand for health care providers in these settings—are associated with

patient characteristics and payment sources.

**GEOGRAPHICAL DETAIL:** National, regional, and state-level analyses are possible using HCUP national databases. For some participating states, a variable indicating the county is also available in the national databases.

**AVAILABILITY:** Many of the HCUP databases are available for purchase through the HCUP Central Distributor ([https://www.hcup-us.ahrq.gov/tech\\_assist/centdist.jsp](https://www.hcup-us.ahrq.gov/tech_assist/centdist.jsp)). Information is also available using the HCUPnet tool (<https://hcupnet.ahrq.gov>), a free online query system based on data from the HCUP. HCUPnet provides access to health statistics and information on hospital inpatient and emergency department utilization.

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

**FEDERAL AGENCY:** Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.ahrq.gov/cahps>

**DESCRIPTION:** The Consumer Assessment of Healthcare Providers and Systems (CAHPS) is a family of surveys overseen by the AHRQ and includes some other surveys used by the Centers for Medicare and Medicaid Services (CMS) that follow CAHPS principles (see <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS>). CAHPS surveys ask patients to report on and assess their health care experiences with health care providers and plans, including hospitals, home health care agencies, doctors, and health and drug plans, among other elements. The topics covered focus on quality of care, such as the communication skills of providers and the accessibility of health care services. Some CAHPS survey results are used as a metric in CMS's Value-Based Purchasing (pay-for-performance) initiatives.

CAHPS surveys are developed with extensive input from stakeholders and are designed to reliably measure patient experiences across a variety of health care settings. Standardized questions and data collection protocols ensure that data can be compared across settings. In many cases, CAHPS measures are statistically adjusted to correct for the large variability in patient types across providers and the various survey modes employed. Detailed survey design information is available at the individual survey level; the CAHPS family of surveys is diverse and each one has a design tailored to its purpose.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The CAHPS survey results can be used by health care providers, health care organizations, insurers, regulators, educators, and other stakeholders to evaluate the quality and patient satisfaction of care provided; to help target interventions to improve quality and satisfaction; and to track improvement. Although the surveys are designed for easy comparison across health care settings and heterogeneous patient populations, the

voluntary nature of the surveys can limit generalizability and specificity across geographic regions.

**GEOGRAPHICAL DETAIL:** The level of detail varies by survey.

**AVAILABILITY:** Data availability is variable depending on whether a survey hails from AHRQ or CMS. Data availability from CMS differs by survey; see <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS>, and visit the page of the specific patient experience survey for more information. Additionally, de-identified research data sets from the CAHPS database may be requested through an application process (see <https://www.ahrq.gov/cahps/cahps-database/data-research>).

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

**FEDERAL AGENCY:** Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <http://meps.ahrq.gov/mepsweb>

**DESCRIPTION:** The Medical Expenditure Panel Survey (MEPS), initiated in 1996, comprises multiple large-scale surveys of individuals and families across the U.S. and their medical providers (e.g., doctors, hospitals, pharmacies, etc.) and employers. In essence, MEPS collects data on the various health care services that Americans use, the frequency of their use, and their associated costs, in addition to specifics on the health insurance plans offered by U.S. employers.

MEPS has several components. The [Household Component](#) provides data from individual households and their members, which is supplemented by data from their medical providers collected in the [Medical Provider Component](#). The [Insurance Component](#) is a separate survey of employers that collects data on employer-provided health insurance. The three MEPS components are described below.

### **Household Component (HC)**

The HC collects information from a sample of individuals and families across the U.S. drawn from a nationally representative subsample of households participating in the [National Health Interview Survey](#) during the previous year. MEPS collects a range of data for all persons in each sampled household, including demographic information, income, employment status, health status, access to and satisfaction with health care, use of medical services, medical costs and sources of payment, and health insurance coverage. Over the past several years, the sample size in the HC has ranged from approximately 12,000 to 14,000 families covering some 30,000 to 35,000 individuals ([https://meps.ahrq.gov/mepsweb/survey\\_comp/hc\\_sample\\_size.jsp](https://meps.ahrq.gov/mepsweb/survey_comp/hc_sample_size.jsp)). The 2010 Full Year Consolidated File from the HC (the year available at the time of this writing) contains 32,846 records.



### **Insurance Component (IC)**

The IC collects information on the private health insurance plans offered by a sample of private- and public-sector employers to their employees. The data collected include employer characteristics; the number and types of insurance plans offered; the eligibility, benefits, and premiums associated with each plan; and information on employer and employee contributions. IC estimates are available on the MEPS website in tabular form for national, regional, state, and metropolitan areas, as well as in publications using IC data and interactive data tools. IC data files are not available for public release. The sample size for the IC has been approximately 42,000 private-sector establishments and 3,000 state and local governments.

### **Medical Provider Component (MPC)**

This component of MEPS collects data from physician offices, hospitals, pharmacies, and other medical providers identified by MEPS HC respondents for the purpose of supplementing or replacing the information obtained in the HC. While dedicated data files containing this supplemental information are not available, the information is fully incorporated into the MEPS HC data files.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The primary focus of the MEPS HC is on health care utilization, expenditures, and insurance coverage and does not include a representative sample of providers for analysis of supply. The data collected on sampled individuals—which include demographic characteristics, health status, health insurance coverage, and use of health services—can be relevant to estimating or predicting demand for health care.

**GEOGRAPHICAL DETAIL:** The level of geographical detail varies by MEPS component. Summary tables are released by the AHRQ for the HC at the national and regional levels and at the state level for selected states (contingent on reliability of state-level estimates). Summary tables for the IC include estimates at the national, state, and metro area levels.

**AVAILABILITY:** Summary tables with findings from MEPS are available at [https://meps.ahrq.gov/mepsweb/data\\_stats/quick\\_tables.jsp](https://meps.ahrq.gov/mepsweb/data_stats/quick_tables.jsp). Downloadable public-use files and documentation for the MEPS IC are available for data analysis at [https://meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files.jsp](https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp). For individual and family-level analyses, researchers and data users with approved research projects can access, at the AHRQ Data Center in Rockville, Maryland, restricted data files covering individuals and families, which otherwise are not publicly released for reasons of confidentiality ([https://meps.ahrq.gov/mepsweb/data\\_stats/onsite\\_datacenter.jsp](https://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp)). For employer-level analyses, researchers with approved projects can access restricted data files covering business establishments through the U.S. Census Research Data Center network (<https://www.census.gov/ces/dataproducts>).

## American Time Use Survey (ATUS)

**FEDERAL AGENCY:** Bureau of Labor Statistics (BLS), U.S. Department of Labor

**WEBSITE:** <https://www.bls.gov/tus>

**DESCRIPTION:** The American Time Use Survey (ATUS), sponsored by the BLS and conducted by the U.S. Census Bureau, provides nationally representative estimates of how Americans spend their time. The survey was developed in the late 1990s and early 2000s as a way to measure the amount of time Americans spend doing both paid work and unpaid, nonmarket work such as unpaid childcare, eldercare, housework, and volunteering. Time spent on leisure, fitness-related, and social activities is also measured.

The ATUS sample consists of individuals who are randomly selected from the population of households who have completed their eighth and final month of interviews for the Current Population Survey (CPS). Selection occurs in three stages. In the first stage, the ATUS sample is obtained by subsampling the CPS sample to ensure that the former is distributed across the states roughly proportionally to the national population represented by each state. In the second stage, households are stratified based on race/ethnicity, the presence and age of children in the household, and the number of adults in households without children. In the third stage, an eligible person (at least 15 years of age) from each household selected in the second stage is randomly chosen as the designated person for the ATUS.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The ATUS, along with additional health-related modules, can be used to describe how workers in different health care occupations spend their time engaging in work-related and non-work-related activities. This information can be used to better understand elements that could contribute to stress and burnout of health care professionals. The ATUS can also be used to understand health-related factors in the general population as indicators of potential demand for certain health care services, such as mental health and nutrition.

**GEOGRAPHICAL DETAIL:** Variables identifying the state in which survey respondents resided at the time of the interview, their Census region, Census division (starting in 2014), and metropolitan status (based on metropolitan statistical area [MSA] definitions from the 2000 or 2010 Decennial Census) are included in the ATUS-CPS data file.

**AVAILABILITY:** ATUS microdata files are freely available for download at <https://www.bls.gov/tus/#data>. For several key pieces of documentation, including the User's Guide, Data Dictionaries, and Coding Lexicons, please visit <https://www.bls.gov/tus/howto.htm>.

## Current Employment Statistics (CES) Survey

**FEDERAL AGENCY:** Bureau of Labor Statistics (BLS), U.S. Department of Labor

**WEBSITES:** <https://www.bls.gov/ces> and <https://www.bls.gov/sae>

**DESCRIPTION:** The Current Employment Statistics (CES) program is a payroll or establishment survey gathering detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls. The CES program serves as a leading economic indicator of current economic trends each month. The program reports employment and payroll information by industry using the North American Industry Classification System (NAICS).

Each month, the CES program surveys some 147,000 businesses and government agencies in the U.S., representing more than 630,000 individual worksites.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The CES program provides information on total employment, payroll costs, and hours worked in health care establishments as grouped by NAICS code. The program provides information on an industry in total and does not collect occupational information. Occupational employment data are collected by the Current Population Survey (CPS) and the Occupational Employment Statistics (OES) program.

**GEOGRAPHICAL DETAIL:** BLS develops estimates of employment statistics at the national, state, and major metropolitan area levels.

**AVAILABILITY:** Tables and data files with national summary information at the industry level are available at <https://www.bls.gov/ces/home.htm#data>. Summary tables and data on state and metro area employment, hours, and earnings are available at <https://www.bls.gov/sae/data.htm>.

## Employment Projections (EP) Program

**FEDERAL AGENCY:** Bureau of Labor Statistics (BLS), U.S. Department of Labor

**WEBSITE:** <http://www.bls.gov/emp>

**DESCRIPTION:** The Employment Projections (EP) program at the BLS provides U.S. labor market projections for the following 10 years. The projections are developed in a stepwise fashion using a series of sequential models and assumptions and are updated biennially. The methodology is described in Chapter 13 of the BLS *Handbook of Methods* (<http://www.bls.gov/opub/hom/pdf/homch13.pdf>; see also [https://www.bls.gov/emp/ep\\_projections\\_methods.htm](https://www.bls.gov/emp/ep_projections_methods.htm)).

The EP program uses survey data from several different sources to develop projections for the labor force, the macro economy, industry output and employment, and occupational employment. Data sources include the Current Population Survey (CPS), the Current Employment Statistics (CES) survey, the Occupational Employment Statistics (OES) survey, and U.S. Census Bureau population projections. Each separate data source has a different sample frame and size.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The EP program includes projections for individual health occupations categorized in the Standard Occupational Classification (SOC) system. Additionally, projections for health occupations by industry or health-related industries by occupation are available (<https://www.bls.gov/emp/#tables>). Career information on the nature of the work, education and training requirements, job outlook, and more can be accessed from the *Occupational Outlook Handbook* at <https://www.bls.gov/ooh/healthcare/home.htm>.

**GEOGRAPHICAL DETAIL:** The projections released by BLS are for the nation as a whole.

**AVAILABILITY:** Detailed projections for 300 industries and 750 occupations, as well as summary tables on selected subjects, are available at <https://www.bls.gov/emp/#tables>. Detailed data files for researchers also are available (<https://www.bls.gov/emp/#data>). The *Occupational Outlook Handbook* (<https://www.bls.gov/ooh>)

also reports projections and other career information for more than 500 detailed occupations.

## Occupational Employment Statistics (OES) Survey

**FEDERAL AGENCY:** Bureau of Labor Statistics (BLS), U.S. Department of Labor

**WEBSITE:** <https://www.bls.gov/oes>

**DESCRIPTION:** The Occupational Employment Statistics (OES) survey is primarily a mail survey that measures occupational employment and wages among full- and part-time wage and salary workers in nonfarm establishments across the nation and, individually, in the 50 states and the District of Columbia, Guam, Puerto Rico, and the Virgin Islands. An “establishment” is defined generally as a discrete physical location at which economic activity occurs (e.g., a factory, restaurant, retail store, hospital, or physician office). The North American Industry Classification System (NAICS) is used to categorize the industry of each establishment.

Respondents report the number of workers in their employ by occupation and across 12 wage ranges. Self-employed persons, owners of and partners in unincorporated firms, unpaid family workers, and household workers are not included in the survey. Occupations are categorized according to the Standard Occupational Classification (SOC) system.

The sample size for reporting estimates of individual occupations varies by occupation, in part depending on the prevalence of the occupation within the establishments sampled. Probability sample panels of about 200,000 establishments are selected semiannually. OES estimates are based on six panels of establishment survey data and rely on data from approximately 1.2 million sampled establishments collected over a three-year period. (See [http://www.bls.gov/oes/current/methods\\_statement.pdf](http://www.bls.gov/oes/current/methods_statement.pdf) for more information.)

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The OES includes information on more than 60 health occupations. Data elements include the estimated total number of individuals employed in an occupation and the mean and percentile values for hourly and annual wages.

Employment and wage information for an occupation can be detailed by industry. Thus, for example, information is available on the employment and wages of



registered nurses nationally in general hospitals, physician offices, home health care, nursing facilities, outpatient centers, and so on (e.g., <https://www.bls.gov/oes/current/oes291141.htm>).

**GEOGRAPHICAL DETAIL:** Data are reported for the nation, by state, and for metropolitan statistical areas, metropolitan divisions, and non-metro areas.

**AVAILABILITY:** Summary tables are available on the OES website. Downloadable Microsoft Excel spreadsheets and text files with occupational employment and wage estimates are also available (see <https://www.bls.gov/oes/tables.htm>).

## American Community Survey (ACS)

**FEDERAL AGENCY:** U.S. Census Bureau, U.S. Department of Commerce

**WEBSITE:** <https://www.census.gov/programs-surveys/acs>

**DESCRIPTION:** The American Community Survey (ACS) is an ongoing statistical survey that samples a small percentage of the population in the U.S. and Puerto Rico annually. The ACS contains a sample of housing units and group quarters and includes information on age, sex, race/ethnicity, language, disability, health insurance status, state of residency and employment, employment status, hours worked, occupation, education, income, household size and characteristics, and family characteristics and relationships, among other items.

There are three types of ACS data files: one-year files, three-year files, and five-year files. The one-year file contains data for a single year, the three-year file combines data from three years, and the five-year file combines five years of ACS data.

The ACS annually collects data from about 2 million households. The sample is based on the U.S. Census Bureau's official inventory of known living quarters, the Master Address File. Data are collected on all individuals in a sampled household and on a sample of individuals in sampled group quarters. About one in 38 households receives an invitation to participate in the ACS each year ([https://factfinder.census.gov/faces/affhelp/jsf/pages/metadata.xhtml?lang=en&type=document&id=document.en.ACS\\_pums\\_csv\\_2015#main\\_content](https://factfinder.census.gov/faces/affhelp/jsf/pages/metadata.xhtml?lang=en&type=document&id=document.en.ACS_pums_csv_2015#main_content)).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The ACS has data elements relevant to analyses of health workforce supply and demand. Variables such as occupation, employment status, location, and hours worked—in addition to demographic information on age, sex, race/ethnicity, etc.—can be used to estimate the size and characteristics of a specific health occupation. Occupation classification coding schemes include the Standard Occupational Classification (SOC) scheme and the Census classification scheme (based upon SOC) (e.g., see <https://www.census.gov/programs-surveys/acs/technical-documentation/code->

[lists.html](#)). The ACS also enables analysis of trends in health workforce supply, given that the ACS is collected continuously and estimates are published annually. Empirical trends derived from the ACS, for example, may be able to track growth in an occupation over time.

Variables related to individual demographics, health insurance status, and disability status may be applicable to studies of health care needs underlying demand for health care providers.

**GEOGRAPHICAL DETAIL:** The geographical detail in estimates published in summary form by the Census Bureau varies based on whether estimates are from a one-year, three-year, or five-year file. The reliability of estimates for geographical units with the smallest population numbers increases as the number of years combined in a file increases. The Census Bureau reports summary estimates for areas with populations of 60,000 or more in one-year files; for areas with populations of 20,000 or more in three-year files; and for areas with fewer than 20,000 persons in five-year files.

In Public-Use Microdata Samples (PUMS), which contain individual household- and person-level data, the smallest geographical units of analysis are areas with a population of at least 100,000 persons.

**AVAILABILITY:** Information is available in summary tabulations produced by the Census Bureau. Microdata files are also available.

The ACS summary data are presented in predefined tabulations of characteristics. The basic unit of analysis is a specific geographic entity—for example, a state or county—for which estimates of persons, families, households, or housing units in particular categories are provided. A user can select specific predefined tabulations through the Census Bureau's American FactFinder (<https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>). Summary data files provide access to the detailed tables through a series of comma-delimited text files on the Census Bureau's file transfer protocol (FTP) site.

With PUMS (microdata) files, in contrast to summary data tabulations and files, it is

the user who determines the structure of the tabulation and the characteristic(s) to be tabulated. PUMS files provide access to ACS microdata for data users in SAS or SPSS format. In the ACS microdata, the basic units are the individual housing units within the housing unit data file and the persons who live in the selected housing units in the person data file. Each record shows most of the information associated with a specific housing unit or individual except for personally identifiable information and items that could be used to identify an individual.

Information on downloading PUMS and summary data files from the ACS is available at <https://census.gov/programs-surveys/acs/data.html>.

Handbooks for ACS data users are available at <https://www.census.gov/programs-surveys/acs/guidance/handbooks.html>.

Details on the ACS survey methodology and content are available at <https://census.gov/programs-surveys/acs/methodology.html>.

## Current Population Survey (CPS)

**FEDERAL AGENCY:** U.S. Census Bureau, U.S. Department of Commerce; Bureau of Labor Statistics (BLS), U.S. Department of Labor

**WEBSITE:** <https://www.census.gov/programs-surveys/cps.html>

**DESCRIPTION:** The Current Population Survey (CPS) is sponsored jointly by the U.S. Census Bureau and the BLS. The CPS is the primary data source used by BLS for the national unemployment rate, among other uses in the analysis of labor force and economic statistics.

The CPS is administered by the Census Bureau using a monthly probability sample of occupied housing units. The actual field work is performed during the calendar week that includes the 19th day of the month, with survey questions asking about activities during the previous week. Households from all 50 states and the District of Columbia are surveyed for four consecutive months; after a subsequent eight-month dormant period, they are surveyed again for another four months before finally leaving the sample. A personal-visit interview is required for all first month-in-sample households. For households in the sample for the second, third, and fourth months, the preferred method of data collection is a telephone interview. A personal interview is again preferred for the household's fifth-month interview (following the sample unit's eight-month dormant period). Fifth-month households are more likely than any other sampled households to be ones in which the previous residents have moved out and been replaced by an entirely different group of residents.

In addition to questions focusing on the labor market, the CPS often includes supplemental questions on related topics such as annual work activity, annual income, contingent employment, worker displacement, job tenure, school enrollment, and veteran status.

During each month of data collection, about 60,000 housing units are eligible for interview, with information collected on all residents 16 years of age or older,

resulting in about 150,000 person-level records in each monthly file (see [http://thedataweb.rm.census.gov/ftp/cps\\_ftp.html](http://thedataweb.rm.census.gov/ftp/cps_ftp.html)).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The CPS has data elements that are relevant to analyses of health workforce supply and labor force participation. Variables such as occupation, employment status, and hours worked—in addition to demographic information such as age, sex, race, and ethnicity—can be used to estimate the size and characteristics of a specific health occupation, contingent on sufficient sample size. Occupation classification codes are provided in the CPS microdata file for the main job and any secondary job a person held during the reference week of the data collection. Currently the codes represent the Census 2010 classification system, which was developed using the *Standard Occupational Classification (SOC) System Manual, 2010*. The CPS also enables the analysis of trends in health workforce supply because it is conducted monthly. The empirical trends derived from the CPS may, for example, illuminate the rate of new entrants into an occupation, as well as exits due to retirement.

**GEOGRAPHICAL DETAIL:** The CPS data files include a variable for the survey participant's state and metropolitan statistical area. There is a variable for county, but most counties are not identified in public-use data files. There are also variables that indicate the metropolitan status of a participant's residence (e.g., metropolitan or nonmetropolitan).

**AVAILABILITY:** Basic monthly CPS data files and the annual March or Annual Social and Economic (ASEC) Supplement data files, along with documentation, can be downloaded from [http://thedataweb.rm.census.gov/ftp/cps\\_ftp.html](http://thedataweb.rm.census.gov/ftp/cps_ftp.html). Summary data tables from the latest ASEC Supplement are available at <https://www.census.gov/programs-surveys/cps/data-detail.html>. Summary data files and tables on labor force statistics are also available from the BLS at <https://www.bls.gov/cps/home.htm>.

More detail on the CPS methodology is available at <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology.html>.

## Economic Census

**FEDERAL AGENCY:** U.S. Census Bureau, U.S. Department of Commerce

**WEBSITE:** <https://www.census.gov/programs-surveys/economic-census.html>

**DESCRIPTION:** The Economic Census (EC) provides comprehensive data on the American economy every five years. Data from the EC can be used to better understand trade and business characteristics, including statistics by industry, location, owner demographics, employment, sales, and performance. Additionally, crucial composite economic indicators (e.g., gross domestic product, production and price indices, and monthly retail sales) rely on the EC for accurate benchmarks. Businesses of all sizes use the information from the EC for strategic planning, including expansion, marketing, and development plans, and for human resources-related metrics, such as the effects of employee benefits. Statistics from the EC can also assist startup businesses in formulating their business plans.

Mandatory EC forms are mailed to more than 4 million businesses in the U.S., across nearly every industry in the private, nonfarm economy, and in every geographic area, including Puerto Rico and other U.S. islands. The majority of businesses with four or more paid employees, and a sample of smaller firms, receive an EC form. This sample is chosen using a stratified sampling procedure with strata based on industry and geography. Basic data for non-selected small businesses are acquired from federal administrative records. Nonprofit organizations, including those in the educational sector, are also incorporated into the EC. Statistics are collected and issued mainly by establishment, which is defined as a business or industrial unit at a single physical location that produces or distributes goods or performs services, regardless of whether it is part of a multi-unit firm.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The EC contains information that describes the characteristics of various health care establishments by industry (e.g., Health Care and Social Assistance), industry subsector (e.g., Ambulatory Health Care Services, Hospitals, Nursing and Residential Care Facilities, Social Assistance), and industry level (e.g., offices of dentists), and by

establishment size. Data include variables such as number of establishments, receipts/revenue, annual payroll, and number of paid employees. While there are limitations to using EC data to characterize the supply of health care professionals, the EC can provide helpful context in broader research and planning initiatives.

**GEOGRAPHICAL DETAIL:** The EC gives statistics at the national and state levels, as well as for metropolitan areas, counties, economic places, and ZIP codes, with reporting that varies from sector to sector. Economic places include both incorporated places and Census Designated Places (CDPs). The most complete range of classifications and statistics are typically available at the national level. As the level of analysis becomes smaller, such as for states and areas within states, there are fewer statistics to avoid revealing information about individual firms.

**AVAILABILITY:** EC statistics are available through American FactFinder at <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>. Additionally, EC tables are downloadable from the Census Bureau's website in XLS, CSV, and PDF file formats at <https://www.census.gov/programs-surveys/economic-census/data/tables.html>.



## **Survey of Business Owners and Self-Employed Persons (SBO)**

**FEDERAL AGENCY:** U.S. Census Bureau, U.S. Department of Commerce

**WEBSITE:** <https://census.gov/programs-surveys/sbo.html>

**DESCRIPTION:** The Survey of Business Owners (SBO) provides comprehensive information on selected economic and demographic characteristics for businesses and business owners. Conducted every five years since 1972, the SBO provides data on gender, ethnicity, race, and veteran status of business owners. Title 13 of the U.S. Code authorizes and mandates this survey. Included are all nonfarm businesses from 20 industries, specifically those filing tax forms as individual proprietorships, partnerships, or corporations of any type and with receipts of \$1,000 or more. Data are collected on a company or firm basis rather than an establishment basis, a company or firm being defined as a business consisting of one or more domestic establishments that the reporting firm specifies as under its ownership or control. The data are accumulated by combining data collected from the SBO with data collected from the national economic census and administrative records. The SBO data sets include all businesses—minority-, nonminority-, and equally minority-/nonminority-owned; female-, male-, and equally female-/male-owned; veteran-, nonveteran-, equally veteran-/nonveteran-owned; and publicly held companies and other businesses whose ownership cannot be classified by the preceding demographic qualifiers. Data are presented by industry classifications and/or geographic area (states, metropolitan and micropolitan statistical areas, counties, and municipalities) and size of firm (employment and receipts).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The SBO contains information that describes the characteristics of various health care businesses by industry (e.g., Health Care and Social Assistance), industry subsector (e.g., Ambulatory Health Care Services, Hospitals, Nursing and Residential Care Facilities, Social Assistance), and industry level (e.g., offices of dentists). Data include estimates of business ownership by gender, ethnicity, race, and veteran status; number of firms with and without paid employees; number of paid employees; and

annual payroll. While there are limitations to using SBO data to characterize the supply of health care professionals, the SBO can provide helpful context in broader research and planning initiatives.

**GEOGRAPHICAL DETAIL:** Data aggregates are offered by states, metropolitan and micropolitan statistical areas, counties, and corporate municipalities, including cities, towns, townships, villages, and boroughs.

**AVAILABILITY:** The SBO data are available on the Census Bureau's website at <https://census.gov/programs-surveys/sbo.html>. The Census Bureau's online, self-service data access tool, American FactFinder (<https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>), allows selective retrieval and downloading of the SBO data.

## Survey of Income and Program Participation (SIPP)

**FEDERAL AGENCY:** U.S. Census Bureau, U.S. Department of Commerce

**WEBSITE:** <https://www.census.gov/sipp>

**DESCRIPTION:** The Survey of Income and Program Participation (SIPP) collects detailed data on the financial situations of households and individuals, including employment status and income, as well as more comprehensive metrics of well-being, such as family dynamics, educational attainment, housing and utility expenses, assets, health insurance, medical expenditures, disability, childcare, food security, and direct or indirect government assistance. This information puts the income and/or benefit reciprocity of individuals and households into a multifactorial social context and allows researchers to study the ways in which these factors exert influence over financial status and movement into or out of assistance programs.

SIPP takes the form of a continuous sequence of national panels, each panel comprising a nationally representative sample interviewed regularly over an approximate four-year period. It serves as an integrated source of data for a range of socioeconomic indicators that all contribute to a nuanced, complex understanding of the nation's economic well-being over time. Most importantly, SIPP data allow for analysis of how the interactions between tax, transfer, and other government and private policies affect households. Policymakers rely on SIPP for information regarding the distribution of income and the effectiveness of government assistance programs. Data are released in cross-sectional, topical modules and longitudinal reports.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** SIPP health-related variables (e.g., health care utilization and medical expenditures, health insurance, disability income) can be analyzed with the SIPP socioeconomic variables (e.g., language spoken at home, educational attainment, employment, demographic characteristics) to provide contextual information on the U.S. population for health workforce planning purposes. The longitudinal SIPP files can be used to examine trends in unemployment, insurance coverage, and health care utilization. This information can be used to broadly compare and identify the need for health

workforce services—for example, health professionals who can provide services in a language other than English, or the need for safety net providers to care for the uninsured.

**GEOGRAPHICAL DETAIL:** SIPP produces national-level estimates for the U.S. population.

**AVAILABILITY:** SIPP data and documentation are public information, as released by the Census Bureau, and downloadable SAS files are available at [http://thedataweb.rm.census.gov/ftp/sipp\\_ftp.html](http://thedataweb.rm.census.gov/ftp/sipp_ftp.html). Additionally, researchers may use the DataFerrett extraction tool (<https://dataferrett.census.gov>) to analyze SIPP data.

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

**FEDERAL AGENCY:** Office of Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/BRFSS>

**DESCRIPTION:** The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project of the CDC and U.S. states and territories. The BRFSS is an ongoing data collection program designed to measure health risk behaviors, preventive health practices, and health care access in the American adult population (18 years of age and older).

The BRFSS was initiated in 1984, with 15 states collecting surveillance data on risk behaviors through monthly telephone interviews. Over time, the number of states participating in the survey increased. By 2001, 50 states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands were participating in the BRFSS. BRFSS field operations are managed by state health departments that follow guidelines provided by the CDC.

Before 2011, the sampling frame for the survey was based on lists of traditional landline telephone numbers. In 2011, 50 states, the District of Columbia, Guam, and Puerto Rico collected samples of both landline and cell phone interviews, while the Virgin Islands collected a sample of landline-only interviews. Information on the methodological changes associated with the 2011 BRFSS and the potential effects of these changes on prevalence estimates is available at [https://www.cdc.gov/brfss/about/brfss\\_today.htm](https://www.cdc.gov/brfss/about/brfss_today.htm).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** BRFSS is one potential source of data on health care utilization and access by the adult population, used to gauge demand for services. It is not intended, in sample design or purpose, for analysis of the supply of providers. The BRFSS data collection enables estimates of health status (including prevalence of chronic conditions), health utilization, and health access.

**GEOGRAPHICAL DETAIL:** State-level estimates and some county- and metropolitan/micropolitan-level estimates are available from the annual BRFSS public-use data file (PUF). Although there is a variable for identifying the county of a respondent, this field is often absent from the PUF to protect the confidentiality of respondents. Hence, information on counties other than those with sizable populations is not available. Nonetheless, there is a variable indicating the metropolitan/micropolitan statistical area status of county areas.

Geographical details for selected cities and counties within metropolitan/micropolitan areas are available in pretabulated form through BRFSS SMART (Selected Metropolitan/Micropolitan Area Risk Trends) at [https://www.cdc.gov/brfss/smart/smart\\_data.htm](https://www.cdc.gov/brfss/smart/smart_data.htm).

**AVAILABILITY:** Annual BRFSS PUFs incorporating the monthly data collected by each state are available from the CDC. Data files and documentation are available at [https://www.cdc.gov/brfss/annual\\_data/annual\\_data.htm](https://www.cdc.gov/brfss/annual_data/annual_data.htm). These data files enable primary analysis of the case-level data.

The BRFSS Web Enabled Analysis Tool (WEAT) enables users to create cross-tabulation analyses and logistic regression analyses from variables available in the WEAT variable lists (see [https://nccd.cdc.gov/s\\_broker/WEATSQL.exe/weat/index.hsql](https://nccd.cdc.gov/s_broker/WEATSQL.exe/weat/index.hsql)). The analysis selected by the user is run through the WEAT system, with the generated results reported on the WEAT website and downloadable in Microsoft Excel spreadsheet form. A suppression protocol is applied—that is, estimates are suppressed when the denominator sample size is less than 50.

As noted, pretabulated data tables, charts, and documentation for selected metropolitan/micropolitan areas are available at [https://www.cdc.gov/brfss/smart/smart\\_data.htm](https://www.cdc.gov/brfss/smart/smart_data.htm). These data files may be used for primary analysis of the case-level data.

## National Ambulatory Medical Care Survey (NAMCS)

**FEDERAL AGENCY:** Division of Health Care Statistics, National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/nchs/ahcd>

**DESCRIPTION:** The National Ambulatory Medical Care Survey (NAMCS) is an annual national survey designed to provide objective, reliable information on the availability and utilization of ambulatory medical care services in the U.S. The survey was conducted annually from 1973 to 1981, in 1985, and annually since 1989.

The NAMCS is based on a sample of practicing physicians and a sample of patient visits to each sampled physician's office-based practice. Each sampled physician is randomly assigned to a one-week reporting period, during which data for a systematic random sample of visits are recorded. Data are obtained on patients' symptoms, physician diagnoses, and medications ordered or provided, and patient demographic characteristics are recorded. In addition, data are intermittently collected on special topics—for example, the 2012 survey included a series of questions on complementary and alternative medicine.

The NAMCS survey platform also incorporates supplemental surveys on various topics given to physicians of certain specialties. Examples include cervical cancer screening (2006–2010), electronic health records/electronic medical records (EHR/EMR) (2008–2013), and physician workflow (2011–2013). The EHR/EMR and physician workflow supplements pertain to the adoption of and benefits and barriers associated with the use of electronic health records in physician office-based practices. Both supplements were sponsored by the DHHS's Office of the National Coordinator for Health Information Technology. A "lookback" module was added in 2012 that collects, on the sampled visit, clinical data on the previous 12 months. This module is triggered when certain cardiovascular diagnoses are made in order to evaluate and monitor the appropriateness of clinical management and outcomes.

Through the 2011 survey year, an annual NAMCS data file typically included about 30,000 (visit) records based on data collected from between 1,200 and 1,400 sampled physicians. As of 2012, the number of office-based physicians in the drawn sample was increased to between 15,000 and 20,000, resulting in a larger visit-level data file. Since 2006, the NAMCS also has sampled providers in community health centers (CHCs), including nurse practitioners, physician assistants, and nurse midwives, as well as physicians. The sample of CHC providers now stands at about 6,000 per year. Only data regarding sampled physicians and their clinical encounters have been included in public-use files (PUFs).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** NAMCS data have been used to make estimates of the volume and type of ambulatory care visits at physician offices. This information may be used to estimate demand for ambulatory care providers by describing the visit rates and patterns across different types of office-based physicians based on patient characteristics and expected sources of payment.

The NAMCS also can be used to make physician estimates when analyzing provider supply. Part of the data collected includes information on physicians and their practices. NAMCS PUFs from 2005 onward contain a physician-level weight that can be used to make estimates of office-based physicians.

Beginning in 2013, the NAMCS is also collecting data on non-physician health personnel in each practice, including nurse practitioners and physician assistants.

**GEOGRAPHICAL DETAIL:** National and regional estimates can be made with sufficient reliability from historical PUFs. The 2012 increase in sample size is intended, in part, to enable estimates for a majority of the states in the U.S.

**AVAILABILITY:** Documentation on NAMCS data files and downloadable PUFs are available from [https://www.cdc.gov/nchs/ahcd/ahcd\\_questionnaires.htm](https://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm). Restricted data files, which include some data elements not available in PUFs, are accessible for approved projects through the Research Data Center at the NCHS (<https://www.cdc.gov/rdc>).



## National Health and Nutrition Examination Survey (NHANES)

**FEDERAL AGENCY:** Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/nchs/nhanes>

**DESCRIPTION:** The National Health and Nutrition Examination Survey (NHANES) is a multifaceted study designed to evaluate the health and nutritional status of U.S. adults and children through a combination of interviews and physical examinations. NHANES is a key program of the National Center for Health Statistics (NCHS), part of the CDC. Beginning in the early 1960s, NHANES was conducted as a series of surveys focusing on different population groups or emerging health issues, such as major diseases, chronic conditions, and health-related behaviors. Since 1999, the survey has been a continuous program, examining a nationally representative sample of approximately 5,000 persons each year without interruption. The current format allows for increased flexibility and timeliness in releasing biannual data sets and estimates on topics of public health interest. The NHANES interview includes demographic, socioeconomic, dietary, and health status questions. The examination piece, conducted by medical personnel, consists of medical, dental, and physiological measures as well as laboratory tests.

The sample for the survey is selected to represent the U.S. population across all ages. However, as a long-running project, sample designs have varied slightly over time; the specific sampling procedures used during a survey year, and subsequent analytic guidelines, are available at <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx>.

Notably, the sample design used since 1999 allows for the production of aggregate-level national estimates from NHANES each year from a multiyear sample design. While annual samples are nationally representative, estimates for single-year data are comparatively unstable, because NHANES can go to only a small number of primary sampling units each year. Furthermore, releasing only one year of data increases the possibility of a breach in confidentiality. These issues, combined with

the analytic limitations of the annual sample, initiated a decision to release data in two-year cycles and to keep the survey content within those years fixed to the extent possible.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** NHANES data provide important indicators of health and nutritional status. While this information cannot be used to describe the health workforce, it can be used at the national level to estimate current and future demand for health care services, such as those associated with mental health and chronic conditions (e.g., obesity, diabetes).

**GEOGRAPHICAL DETAIL:** NHANES is representative at the national level only.

**AVAILABILITY:** All publicly available data and related documentation are available at the NHANES website, <https://wwwn.cdc.gov/nchs/nhanes/default.aspx>. Codebooks and documentation accompany each data set. Restricted data sets may be available for research purposes by applying to the NCHS Research Data Center. For application instructions, visit <https://www.cdc.gov/rdc>.

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

**FEDERAL AGENCY:** Division of Health Care Statistics, National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <http://www.cdc.gov/nchs/ahcd>

**DESCRIPTION:** The National Hospital Ambulatory Medical Care Survey (NHAMCS) collects data on the availability and use of ambulatory care services in hospital emergency and outpatient departments. Data are obtained from a national sample of visits to the emergency departments (EDs), outpatient departments (OPDs), and ambulatory surgery locations (ASLs) of non-institutional general and short-stay hospitals (excluding federal, military, and Veterans Affairs hospitals) in the 50 states and the District of Columbia. Patient visits to these locations are systematically selected over a randomly determined 4-week reporting period.

Historically, approximately 350 to 400 hospitals have participated annually in the NHAMCS, yielding about 30,000 to 35,000 encounter (visit) records annually for ED visits; 30,000 to 35,000 records for OPD visits; and 15,000 to 20,000 records for ASL visits.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NHAMCS is designed to provide estimates of visits to EDs, OPDs, and ASLs; it is not designed to provide estimates of provider supply within EDs, OPDs, and ASLs. The information in the NHAMCS may be used to estimate the use of and demand for services from EDs, OPDs, and ASLs by describing the visit rates and patterns across EDs, OPDs, and ASLs based on patient characteristics and expected sources of payment. Types of visits might be classified by using variables indicating the reason for the visit, procedures during the visit, and patient diagnoses.

**GEOGRAPHICAL DETAIL:** National and regional estimates can be made with sufficient reliability.

**AVAILABILITY:** Documentation on NHAMCS data files and downloadable public-

use files (PUFs) are available at [https://www.cdc.gov/nchs/ahcd/ahcd\\_questionnaires.htm](https://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm). Restricted data files, which include some data elements not available in PUFs, are accessible for approved projects through the Research Data Center at the NCHS (<https://www.cdc.gov/rdc>).

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

**FEDERAL AGENCY:** Division of Health Care Statistics, National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/nchs/nhds>

**DESCRIPTION:** Conducted annually from 1965 to 2010, the National Hospital Discharge Survey (NHDS) was a national probability survey that collected data on the characteristics of inpatients discharged from nonfederal, non-institutional, short-stay hospitals in the U.S. It employed two data collection procedures: manual and automated. In the manual procedure, sample selection and transcription of information from hospital records were performed either by hospital staff or by U.S. Census Bureau staff on behalf of NCHS. The automated procedure entailed the purchase of electronic data files by NCHS from hospitals, commercial organizations, and states. Approximately 45% of respondent hospitals provided data through the latter system.

The NHDS has now been integrated into a new survey, the National Hospital Care Survey (NHCS), along with the emergency department, outpatient department, and ambulatory surgery data collected by the National Hospital Ambulatory Medical Care Survey (NHAMCS).

The sample size of hospitals participating in the NHDS was reduced by half in 2008 due to funding limitations. In 2010, slightly more than 200 hospitals participated in the NHDS, yielding more than 150,000 sampled inpatient discharge records in the data file. In comparison, in 2005, slightly more than 400 hospitals participated, yielding about 375,000 sampled inpatient discharge records.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NHDS focuses on inpatient utilization; it is not intended to provide information on clinician supply in hospitals. The data collected on sampled inpatient discharges—which includes demographic characteristics of patients, expected sources of payment, and use of health services during the inpatient stay—may be relevant to estimate or predict

how the use of and demand for inpatient care (and by inference, inpatient hospital staff) vary by individual characteristics and expected sources of payment.

**GEOGRAPHICAL DETAIL:** Analysis is possible for the nation as a whole and by region using public-use files (PUFs).

**AVAILABILITY:** Downloadable PUFs and documentation are available without charge at [https://www.cdc.gov/nchs/nhds/nhds\\_questionnaires.htm](https://www.cdc.gov/nchs/nhds/nhds_questionnaires.htm). Restricted data files, which include some data elements not available in PUFs, are accessible for approved projects through the Research Data Center at the NCHS (<http://www.cdc.gov/rdc>).

## **National Study of Long-Term Care Providers (NSLTCP)**

**FEDERAL AGENCY:** Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/nchs/nsltcp>

**DESCRIPTION:** The National Study of Long-Term Care Providers (NSLTCP)—which replaces the earlier National Nursing Home Survey, National Home and Hospice Care Survey, and National Survey of Residential Care Facilities conducted by the National Center for Health Statistics (NCHS)—is intended to monitor trends in the supply, delivery, and utilization of paid, regulated long-term care services. The NSLTCP collects primary data from the residential care community and adult day services sectors and incorporates administrative data from the home health agency, nursing home, and hospice sectors. The multifaceted perspective of the survey allows for comparisons among provider sectors both at similar points in time and over time.

Survey methodology varies across the two major components, the Residential Care Community Questionnaire and the Adult Day Services Center Questionnaire (for more detailed information, refer to [https://www.cdc.gov/nchs/nsltcp/nsltcp\\_questionnaires.htm](https://www.cdc.gov/nchs/nsltcp/nsltcp_questionnaires.htm)). The sampling frame for the Residential Care Community Questionnaire is assembled from lists of licensed residential care communities obtained from the licensing agencies themselves in each of the 50 states and the District of Columbia, while that for the Adult Day Services Center Questionnaire is obtained from the National Adult Day Services Association. Administrative data on home health agencies, nursing homes, and hospices are obtained from the Centers for Medicare and Medicaid Services.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSLTCP includes data on the number and types of staff across long-term care providers, in particular the number of full- and part-time registered nurses, licensed practical nurses, certified nurse assistants and related assistants, and social workers. Information on the number and types of residents or patients served is also included.

**GEOGRAPHICAL DETAIL:** Data can be analyzed at the national or state level (for most states).

**AVAILABILITY:** NCHS has conducted two waves of the NSLTCP since 2012. Reports on these waves are available at [https://www.cdc.gov/nchs/nsltcp/nsltcp\\_products.htm](https://www.cdc.gov/nchs/nsltcp/nsltcp_products.htm). The restricted 2012 and 2014 survey data files for adult day services centers and residential care communities are available through the Research Data Center (RDC) at <https://www.cdc.gov/rdc>. These data sets are not public and require the submission of a proposal to the RDC; fees may apply. The data sets of the preceding surveys of the NSLTCP—the National Nursing Home Survey, National Home and Hospice Care Survey, and National Survey of Residential Care Facilities—are publicly available in SAS and ASCII formats at [https://www.cdc.gov/nchs/data\\_access/ftp\\_data.htm](https://www.cdc.gov/nchs/data_access/ftp_data.htm).



## **National Survey of Residential Care Facilities (NSRCF)**

**FEDERAL AGENCY:** Division of Health Care Statistics, National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/nchs/nsrcf>

**DESCRIPTION:** The National Survey of Residential Care Facilities (NSRCF), last fielded by NCHS in 2010, was the first-ever national survey of assisted living and other residential care providers and their residents. The objective was to provide national estimates of these facilities and residents.

Included in the NSRCF are residential care facilities, assisted living residences, board and care homes, congregate care programs, enriched housing programs, homes for the aged, personal care homes, and state-regulated shared housing establishments. To be eligible to participate in the survey, residential care facilities needed to be state regulated, have four or more registered beds, provide room and board with at least two meals per day and around-the-clock onsite supervision, and provide help with personal or health-related daily care. Excluded were facilities licensed to exclusively serve the severely mentally ill or developmentally disabled populations, facilities without any current residents, and nursing homes that did not meet the above criteria.

The NSRCF provides data on 2,302 residential care providers that participated in the survey and on 8,094 residents receiving care from these providers.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSRCF provides data in two separate files. The provider/facility file includes data on provider characteristics, such as physical structure and environment; types of services offered; types of staff employed; benefits for, training of, and job roles of personal care aides; and policies on admission, retention, and discharge. The resident file includes data on resident characteristics, including demographics, involvement in inside and outside activities, use of services, charges for care, health status, and cognitive and physical functioning.

**GEOGRAPHICAL DETAIL:** The primary purpose of the NSRCF is to provide national estimates of the number of residential care facilities operating in the U.S., the number of residents receiving care, and the characteristics of both the facilities and their residents.

**AVAILABILITY:** The public-use data files were released in December 2011, along with a data brief describing first findings and a methods report. Data files can be downloaded from [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NSRCF/2010](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010).

Before downloading the files, data users should review the information at [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF\\_FacilityFileDataReadme.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileDataReadme.pdf) and [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF\\_ResidentsFileDataReadme.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_ResidentsFileDataReadme.pdf).

## National Youth Tobacco Survey (NYTS)

**FEDERAL AGENCY:** Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** [https://www.cdc.gov/tobacco/data\\_statistics/surveys/nyts](https://www.cdc.gov/tobacco/data_statistics/surveys/nyts)

**DESCRIPTION:** The National Youth Tobacco Survey (NYTS) was initiated in 1999 to collect national data on short-, intermediate-, and long-term indicators of tobacco use among American youth that could ultimately inform the design, application, and assessment of relevant tobacco prevention and control programs. The cross-sectional, school-based survey gathers data on tobacco-related knowledge, views, and behaviors in a nationally representative sample of public- and private-school students from grades 6 through 12. The NYTS also serves as a reference point against which to measure success in meeting certain *Healthy People 2020* goals for reducing youth tobacco use.

The survey was designed to produce national estimates at a 95% confidence level with a margin of error of 5% by school level, grade level, sex, and race/ethnicity. Some further evaluations are also supported for subgroups defined by grade level, sex, and race/ethnicity within school-level domains; however, accuracy will vary considerably according to differences in subpopulation sizes.

For the most recently published NYTS (2015), a stratified, three-stage cluster sample design was used to yield a nationally representative sample of U.S. middle-school and high-school students. Sampling procedures involved the selection of counties, groups of small counties, or parts of large counties within each created stratum; selection of schools within the above units; and selection of students within each selected school. Participation was voluntary at both the school and student levels, and participating students remained anonymous. Responses were appropriately weighted to adjust for nonresponse and to ensure proportional consistency with the national population.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** NYTS data can be used to estimate tobacco use among adolescents. This information can be used to plan

tobacco prevention and control programs and to inform other health education programs. The data also can be used to estimate the need for such services as smoking cessation services and care for smoking-related conditions among adolescents and young adults. Data from the NYTS do not include specific health workforce information.

**GEOGRAPHICAL DETAIL:** The NYTS is suitable for national analysis. States can compare their estimates of prevalence of youth tobacco use with national data; however, data from the NYTS is not representative at the state level.

**AVAILABILITY:** NYTS data sets are available for public use; data are available in SAS and Microsoft Access formats. Additionally, PDFs of the questionnaire, codebook, and methodology report from each year are available at [https://www.cdc.gov/tobacco/data\\_statistics/surveys/nyts](https://www.cdc.gov/tobacco/data_statistics/surveys/nyts).

## **Pregnancy Risk Assessment Monitoring System (PRAMS)**

**FEDERAL AGENCY:** Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/prams>

**DESCRIPTION:** The Pregnancy Risk Assessment Monitoring System (PRAMS) is a joint endeavor of the CDC and state health departments. Started in 1987, it is an annual surveillance project to collect state-specific, population-based data on maternal attitudes surrounding pregnancy as well as changes in maternal and child health indicators. Currently, 83% of all U.S. births are described by PRAMS.

Because the PRAMS sample is selected from all women who recently had a live birth, findings are representative of each state's population of women who have recently delivered a live-born infant. Mothers are drawn for participation by states monthly in a stratified systematic sample. Responses are linked to birth certificate data for analysis, which adds demographic and medical information to the data set.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** PRAMS data can be used by researchers and policymakers to design, implement, and measure the progress of programs to improve the health of mothers and infants. The data can also be used to estimate pre- and postnatal utilization of health care services and to gauge the need for additional pre- and postnatal health education programs. Data from PRAMS do not include specific health workforce information.

**GEOGRAPHICAL DETAIL:** Most states and some additional territories participate in PRAMS. Visit <https://www.cdc.gov/prams/states.htm> to view a participation map. Data can be analyzed nationally, at the state level, and in some cases, by county.

**AVAILABILITY:** Publicly available data are provided through PRAMStat (<https://www.cdc.gov/prams/pramstat>), an online data tool that contains more than 250 maternal and child health indicators from PRAMS. The availability of variables by state and year varies. Analytic Research Files are available by formal request. For application instructions and data availability, refer to <https://www.cdc.gov>.

[gov/prams/researchers.htm#data](http://gov/prams/researchers.htm#data). Proposal reviews are conducted once per month. Data can be analyzed using SUDAAN, SAS, SPSS, and STATA.

## Youth Risk Behavior Surveillance System (YRBSS)

**FEDERAL AGENCY:** Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cdc.gov/healthyyouth/data/yrbs>

**DESCRIPTION:** The Youth Risk Behavior Surveillance System (YRBSS) assesses the prevalence of priority health risk behaviors such as violence, unintentionally injurious behaviors, sexual risk behaviors, alcohol and other drug use, tobacco use, unhealthy dietary behaviors, inadequate physical activity, obesity, asthma, and other health-related behaviors among high school students. Data are obtained from representative samples of 9th- through 12th-grade students via national, state, territorial, tribal government, and local school-based surveys administered biennially. The YRBSS used more than 1,700 surveys to collect data from more than 3.8 million high-school students between 1991 and 2015.

For each state, territorial, tribal, and large urban school district, the YRBSS utilizes a two-part, cluster sample design to yield a representative sample of high-school students within its jurisdiction. In the first sampling phase, schools (almost exclusively public) are selected with probability proportional to enrollment. In the second phase, classes of a specific subject or during a specific time-period are randomly selected. In contrast, the national YRBSS uses a three-part, cluster sample design. National YRBSS data are representative of all public- and private-school high-school students in all 50 states and the District of Columbia. It is important to note that the national data sets do not merely aggregate the state YRBSS data sets; they use a separate sample. YRBSS data, both nationally and within states, are weighted (based on student sex, grade level, and race/ethnicity) to adjust for school and student nonresponse and to make the data representative of the population of students from which the sample was drawn.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The YRBSS includes data on adolescent risk behavior including, but not limited to, the use of drugs, alcohol, and tobacco/nicotine products; physical and sexual violence and behaviors; bullying; mental health; obesity and nutrition; and physical activity. These data can be used

to gauge potential demand for health services in the adolescent and young adult populations. Data from the YRBSS do not contain specific health workforce information.

**GEOGRAPHICAL DETAIL:** YRBSS data sets are available at the national level, for most states and some territories, for large urban school districts, and for tribal governments. The availability of a geographical level below national depends on location-specific YRBSS participation, data quality, and data-sharing policies. YRBSS data are not available by ZIP code, census tract, or school for reasons of confidentiality and sample size limitations.

**AVAILABILITY:** National YRBSS data sets and documentation are freely available for download at <https://www.cdc.gov/healthyyouth/data/yrbs/data.htm>. National YRBSS data sets are available in two file formats: ASCII and Microsoft Access. The ASCII data can be converted into SAS and SPSS data sets. State, large urban school district, territory, and tribal government data sets from surveys conducted in 1999 or later are available in SAS, SPSS, ASCII, and Microsoft Access formats; however, data sets from prior to 1999 are available in ASCII format only.



## Health Outcomes Survey (HOS)

**FEDERAL AGENCY:** Centers for Medicare and Medicaid Services (CMS), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS>

**DESCRIPTION:** The Medicare Health Outcomes Survey (HOS) is a patient-reported outcomes metric designed to collect valid, reliable, and clinically significant health status data from the Medicare Advantage (MA) program. Results of the HOS are used to inform quality improvement activities, program oversight, public reporting, pay for performance, and health improvement initiatives for elderly Americans. Originally created to assess physical and mental health outcomes, the HOS continues to be refined to reflect the most recent advances in summarizing these outcomes. Additionally, the HOS is used to collect data for four Healthcare Effectiveness Data and Information Set (HEDIS) effectiveness-of-care parameters: Management of Urinary Incontinence in Older Adults, Physical Activity in Older Adults, Fall Risk Management, and Osteoporosis Testing in Older Women. All managed care organizations with Medicare contracts are mandated to participate. The HOS is administered each year to a random sample of Medicare beneficiaries drawn from participating MA plans. Follow-up surveys are conducted after two years.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The HOS can be used to compare the quality of health plans, assess MA program performance, identify areas for quality improvement, and evaluate improvement over time by identifying trends in the survey data. Researchers can also use the results to monitor the health of the Medicare population and vulnerable subgroups and to estimate the need for health care providers who care for these patient populations.

**GEOGRAPHICAL DETAIL:** Public-use files (PUFs) are available at the de-identified beneficiary level and can be aggregated to the CMS region level. Limited data set (LDS) files are available at the beneficiary level and can be aggregated to the ZIP code, city, county, state, region, and health plan levels; however, health plan

identifiers are masked. Research identifiable files (RIFs) are available at the beneficiary level and can be aggregated to the ZIP code, city, county, state, and Medicare Advantage Organization (MAO) levels for analysis. Cell sizes of 11 or less must be masked for reporting purposes. The HOS PUFs are not intended to be generalizable or to be used for national estimates.

**AVAILABILITY:** HOS data files are available as PUFs, LDS files, and RIFs. HOS PUFs are not generalizable and cannot be used for national estimates. HOS PUFs are created to avoid the identification of any single beneficiary or MAO, and only respondents to the survey are included in the files. HOS PUFs are available at no cost and can be downloaded directly from CMS. For information on how to gain access to HOS non-identifiable data, identifiable data, and limited data set files, please visit <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/FilesForOrderGenInfo>.

## Medicare Claims Data

**FEDERAL AGENCY:** Centers for Medicare and Medicaid Services (CMS), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <http://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html>

**DESCRIPTION:** Medicare claims files contain information collected by Medicare to pay for health care services provided to Medicare beneficiaries. Each data file includes claim-specific procedure and diagnosis information, dates of service, revenue center detail, payment and charge amounts, beneficiary demographic information, and health care provider and facility data. Each Medicare claims file contains information for a single calendar year.

Claims types (file types) are as follows: skilled nursing facility, outpatient, inpatient, hospice, home health agency, durable medical equipment, and carrier (also known as the Physician/Supplier Part B claims file). For more information, see <https://www.resdac.org/cms-data/file-family/Medicare-Claims>.

In addition to the claims data files, which indicate utilization by Medicare beneficiaries, there is the Master Beneficiary Summary File (MBSF), which includes information on beneficiary demographics, entitlement, and enrollment. The MBSF contains data on all Medicare beneficiaries enrolled and/or entitled in a given year. The MBSF includes several segments that can be requested separately, depending on the information needed for a particular project. The available segments are the base segment (which contains data for enrollments in Medicare Parts A, B, and D), the chronic condition segment, the cost and utilization segment, and the national death index segment. The MBSF, in conjunction with claims data, enables analysis of the rate of Medicare utilization per beneficiary and related analyses. (See <https://www.resdac.org/cms-data/files/mbsf> for additional information.)

Medicare claims files contain 100% of the universe of Medicare claims in a year. The exception is the 5% sample Standard Analytical Files (see below).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** Medicare claims data are relevant to assessing health care provider supply and the utilization of and demand for services by the Medicare population. Provider information in claims data for the Medicare population may enable an assessment of the provider supply in an area. The volume of Medicare visits by providers may also be assessed. Coupled with the Medicare enrollment data (i.e., the MBSF), claims data also can be used to assess rates of utilization for different services by individual demographic variables available in the data. Information on supply and utilization, together, may be relevant to assessing the adequacy of supply and access to care for the Medicare population.

**GEOGRAPHICAL DETAIL:** The geographical scope of Medicare claims data is national, with state, county, and ZIP code information available for analyses, contingent on the file type accessed (see below).

**AVAILABILITY:** Medicare claims data are available in research identifiable files (RIFs), limited data set (LDS) files, and, recently, public-use files (PUFs).

RIFs include beneficiary-level protected health information and, thus, requests for RIF data require a Data Use Agreement and review by CMS's Privacy Board to ensure that beneficiaries' privacy is protected and the need for identifiable data is justified. The MBSF is available as a RIF only.

LDS files contain beneficiary-level health information, but selected variables are encrypted, blanked, or ranged. PUFs have been stripped and edited of all information that may be used to identify individuals. In general, PUFs have aggregate-level information on Medicare beneficiary or provider utilization. See the description by the Research Data Assistance Center at <http://www.resdac.org/resconnect/articles/148> for more information about RIFs, LDS files, and PUFs.

RIFs are available as custom files or 5% sample files. For more information on Medicare RIFs and how to request these files, see <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles>.

LDS files also are available in two forms: the 5% sample and the 100% file. See <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets>.

Basic Stand Alone Medicare claims PUFs can be accessed and downloaded directly from the CMS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/BSAPUFS>.

## Medicaid Claims Data (Medicaid Analytic eXtract [MAX] Data)

**FEDERAL AGENCY:** Centers for Medicare and Medicaid Services (CMS), U.S. Department of Health and Human Services (DHHS)

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

**DESCRIPTION:** Medicaid claims-level data are available from CMS's Medicaid Analytic eXtract (MAX) data system (formerly State Medicaid Research Files). The MAX data are extracted from the Medicaid Statistical Information System (MSIS) (<https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis>). The MAX is a set of person-level data files on Medicaid eligibility, service utilization, and payments. Specifically, the MAX includes the personal summary, inpatient, long-term care, pharmacy, and other therapy data sets, and it contains eligibility and utilization records. The Personal Summary File contains one record for every individual enrolled in Medicaid for at least one day during the year. Together, the sets of files include:

- Demographic data (e.g., date of birth, gender, race)
- Basis of eligibility
- Maintenance assistance status
- Monthly enrollment status
- Utilization summary
- Complete inpatient stay records
- Claims for long-term care services provided by nursing facilities, skilled nursing facilities, intermediate care facilities, and independent psychiatric facilities
- Pharmacy claims
- Claim records for all non-institutional Medicaid services, including physician services, labs/X-rays, and clinic services

The MAX development process combines MSIS initial claims, interim claims, voids, and adjustments for a given service into final action events. A MAX file contains information for one calendar year.

Beginning with 2009 data, a file on provider characteristics (MAXPC) has also been made available. Previously, it was not possible to readily conduct provider-based research because the provider identification numbers in MSIS data are largely state specific and undocumented. Starting in February 2009, states were required to include National Provider Identifiers (NPIs) on their MSIS claims. Note that certain classes of nonmedical providers are not required to obtain an NPI. For example, adult day health care, case management, personal care, nonemergency transportation, and many other services are excluded from the NPI requirement. MAXPC, however, does include legacy IDs in addition to the NPI. In addition to information about the provider, MAXPC contains information on the number of different claims (e.g., inpatient, other therapy, long-term care) filed by the provider and the number of beneficiaries for whom claims were filed. MAXPC is intended to be linkable to the MAX data files. For more about MAXPC and the CMS evaluation of the quality and completeness of MAXPC, visit <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXPC.html>.

MAX data files contain 100% of the universe of Medicaid claims.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** Medicaid claims data are relevant to assessing provider supply and health care utilization and demand for the Medicaid population. Provider information in claims data may enable assessment of provider supply in a specific area for the Medicaid population. The volume of Medicaid visits by providers may also be assessed. Coupled with the Medicaid enrollment file (i.e., personal summary data set), claims data also can be used to assess rates of utilization for different services across individual demographic variables in the data. Information on supply and utilization, together, may be relevant to assessing the adequacy of supply and access to care for the Medicaid population.

**GEOGRAPHICAL DETAIL:** The geographical scope of the MAX files is national, with state, county, and ZIP code information available for analysis.

**AVAILABILITY:** Because the MAX data contain individually identifiable data, availability is restricted by the Privacy Act. MAX data files are available for approved research activities only through a Data Use Agreement (DUA) with CMS. Only approved academic research projects and certain government agencies are entitled to a DUA to obtain MAX data. The cost of MAX data is dependent on the number of states, years, and file types requested. For help in preparing data requests (including DUAs), contact the Research Data Assistance Center (<https://www.resdac.org/cms-data/file-family/Medicaid-Analytic-Extracts-MAX>).

MAXPC is downloadable directly from the CMS website, without cost, at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXPC.html>.



## Medicare Current Beneficiary Survey (MCBS)

**FEDERAL AGENCY:** Centers for Medicare and Medicaid Services (CMS), U.S. Department of Health and Human Services (DHHS)

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

**DESCRIPTION:** The Medicare Current Beneficiary Survey (MCBS), sponsored by the CMS's Office of Enterprise Data and Analytics (OEDA) in partnership with the Center for Medicare and Medicaid Innovation (CMMI) through a contract with NORC at the University of Chicago, is an ongoing, multipurpose, longitudinal survey of a representative sample of the national Medicare population. It has been administered continuously for more than 25 years, encompassing more than 1 million interviews covering the population of beneficiaries in the U.S., the District of Columbia, and Puerto Rico. The MCBS is designed to aid CMS in directing, monitoring, and assessing Medicare programs by determining expenditures, coverages, and sources of payment for all services used by Medicare beneficiaries, including copayments, deductibles, and noncovered services. Additionally, the MCBS follows outcomes over time, such as changes in health status, Medicaid spenddowns, and the impacts of changes within the Medicare program on satisfaction with care and how care is received.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The MCBS focuses on access to care, health care utilization, and costs for elderly persons (i.e., Medicare beneficiaries). The sampling framework of the MCBS is not intended to provide a representative sample of providers for use in an analysis of provider supply, although the MCBS can provide information on the types of providers used by senior citizens and the volume of health utilization across different providers. The detailed data in the MCBS allow for estimating or predicting how demand for health care—and, by extension, for health care providers—among the elderly population varies by demographic characteristics and health status. The MCBS includes beneficiaries' evaluation of their access to and satisfaction with health services, data applicable to assessing the adequacy of health care and the health workforce.

**GEOGRAPHICAL DETAIL:** The public-use data file (PUF) does not contain geographical location information, but files available via formal request include a variable indicating the U.S. Census Bureau geographic division or region.

**AVAILABILITY:** A PUF and accompanying documentation for the 2013 MCBS is available free for download on the MCBS PUF webpage. The PUF is user friendly, with select data items that assist researchers in conducting analyses on health disparities, access to and satisfaction with health care, and the health status of Medicare beneficiaries. Data that are more detailed are available for purchase directly from CMS. Access to these additional data requires a Data Use Agreement and must be requested through the Limited Data Set File Process found at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets>.

## National Provider Identifier (NPI) File

**FEDERAL AGENCY:** Centers for Medicare and Medicaid Services (CMS), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand>

**DESCRIPTION:** The Health Insurance Portability and Accountability Act (HIPAA) requires HIPAA-covered health care providers to obtain and use a National Provider Identifier (NPI), which is a unique identification number. Under HIPAA, covered health care providers are those who transmit health information in electronic form in connection with a transaction for which the Secretary of Health and Human Services has adopted a standard. HIPAA-covered providers include individuals (e.g., physicians, nurses, dentists, chiropractors, physical therapists, and pharmacists) and organizations (e.g., hospitals, home health agencies, clinics, nursing homes, residential treatment centers, laboratories, ambulance companies, group practices, health maintenance organizations, suppliers of durable medical equipment, and pharmacies). However, certain classes of nonmedical providers are not required to obtain an NPI, including those involved in case management, personal care, nonemergency transportation, and other services. Individual providers and covered organizations register and obtain an NPI through the National Plan and Provider Enumeration System (NPPES) administered by CMS (<https://nppes.cms.hhs.gov>).

The NPI data file contains a record for each individual provider and health care organization with an NPI number. The full replacement data file is updated monthly. Data elements in the file include the NPI number, entity type (individual or organization), provider name, business practice location address, provider taxonomy (which includes provider type and specialty differentiation), last update date, and gender, among other data fields. The public data file does not contain social security numbers, Internal Revenue Service individual taxpayer identification numbers, or dates of birth. Additionally, CMS releases a full replacement monthly NPI deactivation file with deactivated NPIs and the deactivation dates in Microsoft Excel format.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NPI file can be a data source for enumerating the supply of those providers in the health workforce who are defined as covered providers under HIPAA.

**GEOGRAPHICAL DETAIL:** The detailed business practice location addresses in the NPI file, which include the ZIP code, offer the opportunity for geographical coding to small areas.

**AVAILABILITY:** The full public-use data file, which is updated monthly, can be downloaded at [http://download.cms.gov/nppes/NPI\\_Files.html](http://download.cms.gov/nppes/NPI_Files.html). Weekly update files and a monthly deactivation file are also available for downloading from the website.

Before downloading, users should review the information at <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination.html>.

## National Survey of Children's Health (NSCH)

**ORGANIZATION:** Data Resource Center for Child and Adolescent Health (DRC)<sup>1</sup>

**WEBSITE:** <http://www.childhealthdata.org/learn/NSCH>

**DESCRIPTION:** The National Survey of Children's Health (NSCH) provides detailed data on children in the U.S. up to 17 years of age. Conducted three times between 2003 and 2012, the survey covers numerous interconnected aspects of children's lives, including their physical and mental health, access to health care, and family, neighborhood, educational, and social environments. The survey was conducted again by the U.S. Census Bureau in 2016, with new data expected in 2017. Notably, the National Survey of Children with Special Health Care Needs (see next entry) has been integrated into the latest version. These data are crucial to understanding the health status and health services needs of children in the U.S. The results of the NSCH can inform the design of state- and national-level policies and programs to improve child and family health outcomes.

The NSCH collects information from a random sample of U.S. households and includes a brief initial survey to determine if children are present in the household, and if so, the number of children, as well as the age, sex, and special health care needs of each child. A portion of those parents or guardians indicating the presence of one or more children in the household then receive a second, lengthier survey about the health of one of the children in the household. The survey was previously conducted by telephone and shifted to an online format for 2016. When sampling weights are used, the subsequent assessments are representative of all noninstitutionalized children up to 17 years of age both in the U.S. as a whole and within each state. These estimates do not generalize to the parent population, however, and researchers are advised to avoid statements about parental statistics based on the NSCH.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSCH contains data on physical and mental health conditions, as well as disabilities, for children up to 17

---

<sup>1</sup> The DRC is funded by a cooperative agreement with HRSA's Maternal and Child Health Bureau.

years of age. The NSCH also contains information on the use of health care services, including whether, where, and how often the child goes for health care visits; accessibility of health care services; experience with the child's health care providers and care coordination; health insurance coverage and cost of the child's health care; and demographic, educational, and socioeconomic data. These data can be used in health workforce planning efforts to better understand children's health care needs and utilization patterns and estimate the need for pediatric health care services and providers.

**GEOGRAPHICAL DETAIL:** Data are available at the national, state, and metropolitan levels. Public-use files (PUFs) for the surveys have state identifiers and a binomial variable distinguishing children living in metropolitan statistical areas (MSA) with more than 500,000 persons. This indicator is suppressed when the population for the MSA or non-MSA area is less than 500,000. ZIP codes are collected by the National Center for Health Statistics (NCHS); however, these data are not distributed in the PUF for confidentiality reasons. Interested investigators may submit a proposal to analyze ZIP code data onsite at the NCHS's Research Data Center.

**AVAILABILITY:** Categorized and cleaned data sets are available from the Data Resource Center for Child and Adolescent Health. To request a data set in SAS or SPSS format, visit <http://www.childhealthdata.org/help/dataset>. All data are free, although some exceptions may apply to for-profit organizations. The original SAS version of the data set (without constructed indicators) can also be retrieved at <https://www.cdc.gov/nchs>.

## National Survey of Children with Special Health Care Needs (NS-CSHCN)

**ORGANIZATION:** Data Resource Center for Child and Adolescent Health (DRC)<sup>2</sup>

**WEBSITE:** <http://www.childhealthdata.org/learn/NS-CSHCN>

**DESCRIPTION:** The National Survey of Children with Special Health Care Needs (NS-CSHCN) was designed to examine the health and functional status of children with special health care needs in the U.S. The survey was conducted three times between 2001 and 2010, and has been integrated into the National Survey of Children's Health as of 2016. The NS-CSHCN assessed the incidence and impact of special health care needs among children under 18 years of age both nationally and within states. The survey aimed to create a portrait of this population's physical, emotional, and behavioral health and also investigated the extent to which these children had suitable health insurance, access to services, and medical homes. Additionally, information was collected on functional difficulties, care coordination, satisfaction with care, transition services, and the impact of chronic conditions on the child's family.

Sample design, eligibility, survey questions, and methodology varied for the various survey years (2003, 2007, 2011–2012, and 2016–2017). All versions of the survey were weighted to be representative of the U.S. population of noninstitutionalized children up to 17 years of age. Design and Operations Manuals created by the National Center for Health Statistics (NCHS) offer detailed information on the NS-CSHCN survey sampling and data collection procedures. These are available at <http://www.childhealthdata.org/learn/NS-CSHCN/resources/methods>.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NS-CSHCN contains data on physical and mental health conditions, as well as disabilities, for children up to 17 years of age who have special health care needs. The NSCH also contains information on the use of health care services, including whether, where, and how often the child goes for health care visits; accessibility of health care services;

---

<sup>2</sup> The DRC is funded by a cooperative agreement with HRSA's Maternal and Child Health Bureau.

experience with the child's health care providers and care coordination; transition plans from pediatric to adult care; health insurance coverage and cost of the child's health care; and demographic, educational, and socioeconomic data. These data can be used in health workforce planning efforts to better understand the health care needs and utilization patterns of children with special health care needs and estimate the need for appropriate pediatric health care services and providers.

**GEOGRAPHICAL DETAIL:** Survey data are available at the national and state levels. National and state data can be refined to evaluate differences by race/ethnicity, income, type of health insurance, and other demographic and health status characteristics.

**AVAILABILITY:** The DRC's NS-CSHCN data set includes all variables from the NSCH public-use data files and can be requested at <http://www.childhealthdata.org/help/dataset>. All indicators in the data set can also be viewed using the DRC interactive data query. The "details" link on each DRC data query page provides the numerator, denominator, revisions, methods notes, treatment of unknown variables, and history with respect to each variable. Additionally, available SAS and SPSS codebooks provide documentation and the SAS/SPSS program statements used in the development of variables and subgroups featured in the interactive data query.

A presentation from the NCHS that provides examples of SAS, SUDAAN, and STATA code for obtaining NS-CSHCN estimates of variability and statistical significance is available at <http://www.childhealthdata.org/docs/nsch-docs/nsch-presentation-with-sas-suddan-spss-examples-pdf>.



## HRSA Data Warehouse

**FEDERAL AGENCY:** Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://datawarehouse.hrsa.gov/default.aspx>

**DESCRIPTION:** The HRSA Data Warehouse (HDW) provides maps, data, reports and dashboards to the public. The data integrates with external sources, such as the U.S. Census Bureau, providing information about HRSA's grants, loan and scholarship programs, health centers and other public health programs and services. Several HDW data sources are listed in the following entries: the Area Health Resources Files, Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps), National Sample Survey of Registered Nurses (NSSRN), and the National Sample Survey of Nurse Practitioners (NSSNP). However, the HDW contains many more data sources on their website, which may be of use to health workforce researchers. Visit their website to learn more.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** There are many purposes associated with the analysis of health workforce data most importantly it is essential for forecasting supply and demand needs, planning educational programs, developing regulatory policies, and evaluating the impact of policy decisions. The available data in the HRSA data warehouse helps our nation prepare and build a workforce that will improve public health by expanding access to care, achieving health equity and providing high quality care to all Americans.

**GEOGRAPHICAL DETAIL:** Data are available at the national, state, and county levels.

**AVAILABILITY:** The HRSA Data Warehouse allows users to interact with data in charts, tables, reports. The HDW manages data from more than 30 HRSA and other federal government sources. (<https://datawarehouse.hrsa.gov/default.aspx>).

## Area Health Resources Files (AHRF)

**FEDERAL AGENCY:** National Center for Health Workforce Analysis (NCHWA), Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <http://arf.hrsa.gov>

**DESCRIPTION:** The Area Health Resources Files (AHRF) contain county-level data on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. There are more than 6,000 variables, with one data record for each county in the U.S. The data in the AHRF are a compilation of the data available from multiple data sources, including data provided for the AHRF by the American Medical Association, the American Hospital Association, the American Dental Association, the U.S. Census Bureau, the Centers for Medicare and Medicaid Services, and the National Center for Health Statistics, among others. The AHRF is released annually.

The AHRF entails no sampling, as it contains data on all of the counties in the U.S., which number more than 3,000.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The AHRF contains data on a range of providers to enumerate workforce supply in a county, including physicians (by specialty and in total), dentists, optometrists, pharmacists, nurses, physician assistants, occupational therapists, and physical therapists, among others.

**GEOGRAPHICAL DETAIL:** Analysis can be performed at the county level or aggregated to the state or national level. County and state identification variables (e.g., Federal Information Processing Standard [FIPS] state and county codes) enable linking AHRF data to other data sets.

The AHRF also contains variables that describe the geographical characteristics of a county, including a metropolitan/micropolitan/non-Core-Based Statistical Area indicator code; a field that indicates if a county of a metropolitan or micropolitan

statistical area is either central or outlying; population density; land area; and typology codes developed in the mid-2000s by the U.S. Department of Agriculture indicating such property characteristics as rural/urban continuum, urban influence, farming dependence, mining dependence, manufacturing dependence, or persistent poverty.

**AVAILABILITY:** AHRF data can be downloaded at no cost, or a CD can be purchased for a nominal fee. Information and a download link can be found at <https://datawarehouse.hrsa.gov/topics/ahrf.aspx>. The direct download site is [https://datawarehouse.hrsa.gov/data/datadownload.aspx#MainContent\\_ctl00\\_gvD D\\_lbl\\_dd\\_topic\\_ttl\\_0](https://datawarehouse.hrsa.gov/data/datadownload.aspx#MainContent_ctl00_gvD D_lbl_dd_topic_ttl_0).

## Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps)

**FEDERAL AGENCY:** Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://bhw.hrsa.gov/shortage-designation>

**DESCRIPTION:** HRSA shortage designation criteria are used to determine whether a given geographic area, special population, or facility qualifies as a Health Professional Shortage Area (HPSA) or a Medically Underserved Area/Population (MUA/P). HPSAs are shortages of primary medical care, dental, or mental health providers, while MUA/Ps are only shortages of primary care providers. Special population HPSAs and MUPs are groups of people for whom health care availability is limited due to insurance status or economic, cultural, or linguistic barriers, among others. Facility designations are limited to HPSAs.

Both types of designations require a rational service area (RSA), which assesses the population (or special population) within a geographic area. Under HPSA guidelines, the proposed RSA must meet the population-to-provider ratio, and services in contiguous areas must be inaccessible to the population of the RSA due to socioeconomic factors, excessive distance, current HPSA status, or overutilization of services. By law, HPSAs require periodic redesignation.

MUA/P designations are based on an index value that considers infant mortality rate, poverty rate, percentage of elderly, and the population-to-primary-care-physician ratio for the area or population. MUA/Ps do not need redesignation and do not require an analysis of contiguous areas.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** HPSAs and MUA/Ps are eligible to receive certain federal resources, including scholarship and loan repayment programs, bonus payments, Federally Qualified Health Center status, and enhanced reimbursement to Rural Health Clinics. Benefits vary by shortage area type.

Both HPSAs and MUA/Ps programs are designed to help communities recruit and retain primary care medical providers (e.g., physicians, nurse practitioners, physician assistants), nurses, dental providers, and mental and behavioral health care providers (e.g., psychiatrists, health service psychologists, licensed clinical social workers, licensed professional counselors, marriage and family therapists, psychiatric nurse specialists, and mental health nurse practitioners and physician assistants).

HPSAs designations are used to recruit health care professionals through HRSA's National Health Service Corps loan repayment and scholarship programs, while both HPSAs and MUA/Ps can recruit physicians through state and Federal J-1 visa waiver programs. A number of states have specific service obligated programs that may use HPSA and/or MUA/P designations for placement of providers.

**GEOGRAPHICAL DETAIL:** RSAs are single counties, groups of counties, groups of townships in rural areas, or groups of census tracts in urban areas in which inadequate health services are available to residents or a special population within the RSA.

**AVAILABILITY:** Federal shortage areas are available via the HRSA Data Warehouse, which allows users to interact with data in charts, tables, reports, and maps (<https://datawarehouse.hrsa.gov/topics/shortageAreas.aspx>).

## National Sample Survey of Registered Nurses (NSSRN)

**FEDERAL AGENCY:** Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSRN.aspx>

**DESCRIPTION:** First conducted in 1977, HRSA's Bureau of Health Workforce has conducted the National Sample Survey of Registered Nurses (NSSRN) approximately every four years through 2008. The data from the NSSRN provide the opportunity to assess trends in and make projections on the nursing workforce. The survey asks registered nurse respondents about various aspects of their professional experience and current status, including their education and training, professional nursing certifications, current and recent employment, annual income, and the states in which they are licensed. It also obtains demographic information.

From 1977 to 2004, a nested sampling frame was employed, which attempted to select nurses from all states with equal probability. In 2008, the survey was redesigned to allow for stratified systematic sampling in each state, with samples selected from current state licensure lists and multiple strata for age, licensure, and employment commuting. In its present form, the NSSRN applies a state-specific sampling weight to each nurse's record, with adjustments for those nurses with multiple state licenses. Although some nurses may be sampled in sequential editions of the survey, the NSSRN is purely a cross-sectional survey and does not aim to provide longitudinal data.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSSRN provides information on the supply and characteristics of registered nurses in the U.S. This information can be used to assess trends in supply, diversity, educational attainment, age structure, and other characteristics to describe current supply and forecast future supply of registered nurses.

**GEOGRAPHICAL DETAIL:** Data are available at the national, state, and county levels.

**AVAILABILITY:** For each survey year, HRSA has prepared two public-use data files (PUFs) in ASCII file format. For surveys conducted in 2008, SAS and SPSS data files are also available. PUFs and documentation for any of the NSSRN surveys may be downloaded at <https://datawarehouse.hrsa.gov/data/dataDownload/nssrnDownload.aspx>.

## National Sample Survey of Nurse Practitioners (NSSNP)

**FEDERAL AGENCY:** Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://datawarehouse.hrsa.gov/data/dataDownload/aboutNSSNP2012.aspx>

**DESCRIPTION:** The National Sample Survey of Nurse Practitioners (NSSNP) is conducted every four years to better understand the supply, distribution, and role of NPs in the health care system. NPs have become increasingly relied upon in the U.S. to provide high-quality primary care, particularly as demand for services has risen. In 2012, HRSA administered a nationally representative sample survey that gathered detailed information on NP demographics, education and training, employment, and practice patterns.

For the 2012 survey, HRSA consulted state licensing boards to obtain lists of all actively licensed NPs in the U.S. Probability matching was used to identify and eliminate duplicate records and create a single national sampling frame, from which a sample of NPs was selected with probability proportional to size. Sample weights were created to balance survey design and nonresponse, and data were cleaned for errors.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSSNP provides information on the size and characteristics of the NP workforce, including specialty, employment setting, job satisfaction, services provided, and billing and practice arrangements. This information can be used to better understand the NP workforce and provide information to support health workforce policy and education decisions.

**GEOGRAPHICAL DETAIL:** The data obtained from the survey are representative at the national level and at the state level for some larger states.

**AVAILABILITY:** Public-use files and documentation for the 2012 NSSNP may be downloaded at <https://datawarehouse.hrsa.gov/data/dataDownload/nssnp2012Download.aspx>. A restricted data file from the 2012 NSSNP with state, ZIP code, and other personal information is available through an application process



maintained by the Research Data Center at the National Center for Health Statistics. These data are available in SAS and ASCII formats.

## **Integrated Postsecondary Education Data System (IPEDS)**

**FEDERAL AGENCY:** National Center for Education Statistics (NCES), U.S.  
Department of Education

**WEBSITE:** <https://nces.ed.gov/ipeds>

**DESCRIPTION:** The Integrated Postsecondary Education Data System (IPEDS) comprises a series of interrelated surveys administered annually by the NCES, part of the U.S. Department of Education. IPEDS collects institutional-level data from postsecondary institutions in the U.S. (50 states and the District of Columbia) and other U.S. jurisdictions. Based on Section 490 of the Higher Education Amendments of 1992 (P.L. 102–325), IPEDS is mandatory for any institution that participates in or is applying for participation in any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended (20 USC 1094[a][17]). Non–Title IV institutions can voluntarily respond to IPEDS.

IPEDS data are collected on the following topics and stored in separate data files: institutional characteristics, enrollments, completions (number of degrees and other formal awards conferred), graduation rates, employees and salaries, financial information, and student financial aid. Data files are released annually.

IPEDS data are not based upon a sample but, rather, represent Title IV and non–Title IV institutions as described above.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** IPEDS can be used to identify and enumerate postsecondary educational institutions with programs that lead to specific occupations in health care and enumerate graduates in those occupations. For example, the classification of an instructional program called “Health Professionals and Related Programs” includes “instructional programs that prepare individuals to practice as licensed professionals and assistants in the health care professions and related clinical sciences and administrative and support services.” In general, the completions component of the IPEDS survey collects data on the number of degrees and certificates conferred in postsecondary education programs by level of degree (associate’s, bachelor’s, master’s, and doctor’s) and by

length of program for sub-baccalaureate and post-baccalaureate certificates. Data are collected on the race, ethnicity, and gender of recipients and their programs of study. The data from this component reflect all formal awards (i.e., degrees, diplomas, certificates) conferred for a given reporting year.

**GEOGRAPHICAL DETAIL:** The addresses of the postsecondary educational institutions in the institution file include street and number, city, state, and ZIP code, which offer the opportunity for analysis at the national, state, or local level.

**AVAILABILITY:** IPEDS data are available to researchers and others through the IPEDS data center at <https://nces.ed.gov/ipeds/datacenter>.

## National Mental Health Services Survey (N-MHSS)

**FEDERAL AGENCY:** Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.samhsa.gov/data/mental-health-facilities-data-nmhss>

**DESCRIPTION:** The National Mental Health Services Survey (N-MHSS), directed by SAMHSA, is an annual survey of all publicly and privately operated specialty mental health care facilities in the U.S. designed to collect a range of data on these facilities and the patients they serve. The survey covers public and private psychiatric hospitals; nonfederal general hospitals with separate psychiatric units; Veterans Affairs medical centers; residential treatment centers for children and for adults; outpatient, day treatment, and partial hospitalization mental health facilities; and multisetting or nonhospital mental health facilities. Additionally, every other year (beginning in 2014), the survey collects statistical information on the numbers and demographic characteristics of persons served in these treatment facilities. Treatment facility information from the N-MHSS is used to inform SAMHSA's Behavioral Health Treatment Services Locator (<https://findtreatment.samhsa.gov>).

Full reports detailing the survey frame, questionnaire, data collection efforts, and results of the N-MHSS are available for each year from 2010 on at <https://www.samhsa.gov/data/mental-health-facilities-data-nmhss/reports>. Select the row for the year of interest to access the corresponding report(s).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** While the N-MHSS cannot be used to describe the behavioral health workforce, it can be used to describe mental health facilities, the services they provide, payment mechanisms, and the patient populations they serve. This information may be valuable in answering questions about access to mental health services and utilization, conducting comparative analyses, and forecasting future mental health treatment resource requirements.

**GEOGRAPHICAL DETAIL:** Data have been analyzed at the national, Census region, and state levels.

**AVAILABILITY:** Data are available only as analyzed through SAMHSA reports found on the N-MHSS website.

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

**FEDERAL AGENCY:** Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.samhsa.gov/data/substance-abuse-facilities-data-nssats>

**DESCRIPTION:** The National Survey of Substance Abuse Treatment Services (N-SSATS) is an annual survey of all public and private substance abuse treatment facilities in the U.S. It is designed to collect multipurpose data that can be used to evaluate the nature and scope of services provided and thus assist SAMHSA, state and local governments, and other stakeholders in predicting treatment resource requirements and analyzing trends in treatment.

Full reports detailing the survey frame, questionnaire, data collection efforts, and results of the N-SSATS are available for each year from 2010 on at <https://www.samhsa.gov/data/substance-abuse-facilities-data-nssats/reports>. Select the row for the year of interest to access the corresponding report(s).

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** While the N-SSATS cannot be used to describe the substance abuse treatment workforce, it can be used to describe substance abuse treatment facilities, the services they provide (including opioid treatment programs), and client counts (e.g., counts of clients by service type and number of beds designated for treatment). This information may be valuable in answering questions about access to substance abuse treatment services and utilization, conducting comparative analyses, and forecasting future substance abuse treatment resource requirements.

**GEOGRAPHICAL DETAIL:** Data have been analyzed at the national and state levels.

**AVAILABILITY:** Data are available only as analyzed through SAMHSA reports found on the N-SSATS website.

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

**FEDERAL AGENCY:** Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (DHHS)

**WEBSITE:** <https://www.samhsa.gov/data/population-data-nsduh>

**DESCRIPTION:** The National Survey on Drug Use and Health (NSDUH; known prior to 2002 as the National Household Survey on Drug Abuse), conducted by SAMHSA, collects data on the use of illegal drugs, alcohol, and tobacco by noninstitutionalized U.S. residents aged 12 years and older. The NSDUH also gathers data on mental disorders, their co-occurrence with substance use, and the management of both problems. The data provided by the NSDUH can be used to evaluate the incidence of substance use and/or mental health conditions, and the need for treatment services, among specific demographic or geographic subgroups.

Methodological resource books and survey results are available at <https://www.samhsa.gov/data/population-data-nsduh/reports>.

**RELEVANCE FOR HEALTH WORKFORCE ANALYSIS:** The NSDUH estimates the prevalence of drug, alcohol, and tobacco use, as well as mental health disorders and conditions. This type of information may be used to estimate current and future demand for behavioral health and substance abuse treatment professionals at different geographic levels.

**GEOGRAPHICAL DETAIL:** Data are available at the national, state, and some substate levels.

**AVAILABILITY:** Data are available only as analyzed through SAMHSA reports found on the NSDUH website.

## Appendix

### Glossary of Acronyms

Acronym	Definition
ASCII	American Standard Code for Information Interchange; a character encoding standard for electronic communication.
SAS	A software package for advanced analytics, multivariate analyses, data management, and predictive modeling.
SPSS	A statistical software package for advance analytics, predictive modeling, and data manipulation.
STATA	Data analysis and statistical software for data management, statistical analysis, graphics, simulations, regression, and custom programming.
SUDAAN	A statistical software package for the analysis of complex sample survey data.