

Compendium of Federal Data Sources to Support Health Workforce Analysis

2022 Edition



Compendium of Federal Data Sources to Support Health Workforce Analysis

2022 Edition

December 2022



Health Resources & Services Administration

National Center for Health Workforce Analysis
Health Resources and Services Administration
5600 Fishers Lane
Rockville, MD 20857

Web: www.hrsa.gov



Center for Health Workforce Studies
School of Public Health, University at Albany
State University of New York
1 University Place, Suite 220
Rensselaer, NY 12144-3445

Phone: (518) 402-0250
Web: www.healthworkforceta.org
Email: info@healthworkforceta.org

PREFACE

This compendium provides a summary of 45 federal data sources that can be used to support health workforce analysis. The compendium was prepared by the Health Resources and Services Administration's National Center for Health Workforce Analysis (NCHWA) and the Health Workforce Technical Assistance Center (HWTAC).

NCHWA 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 NCHWA, visit <https://data.hrsa.gov/topics/health-workforce/data-research>.

HWTAC was established in 2013 to support the efforts of NCHWA and it provides technical assistance to states and organizations that engage in health workforce planning. HWTAC conducts several projects each year designed to provide expert assistance with health workforce data collection, analysis, and dissemination. HWTAC is based at the Center for Health Workforce Studies (CHWS) at the School of Public Health, University at Albany, State University of New York.

HWTAC is supported by the Health Resources and Services Administration (HRSA) of the US Department of Health and Human Services (HHS) as part of an award totaling \$450,000 with 0% financed with nongovernmental sources. The contents in this compendium are those of the authors and do not necessarily represent the official views of, nor an endorsement, by HRSA, HHS, or the US Government. For more information, visit <https://www.hrsa.gov>.

December 2022

SUGGESTED CITATION:

Compendium of Federal Data Sources to Support Health Workforce Analysis, 2022 Edition. Rensselaer, NY: Health Workforce Technical Assistance Center, Center for Health Workforce Studies, School of Public Health, SUNY Albany, National Center for Health Workforce Analysis, Health Resources and Services Administration; December 2022.

TABLE OF CONTENTS

Introduction	2
Agency for Healthcare Research and Quality (AHRQ)	
Consumer Assessment of Healthcare Providers and Systems (CAHPS).....	4
Healthcare Cost and Utilization Project (HCUP)	5
Medical Expenditure Panel Survey (MEPS).....	6
Bureau of Labor Statistics (BLS)	
American Time Use Survey (ATUS)	9
Current Employment Statistics (CES) Survey	10
Employment Projections (EP) Program	11
Occupational Employment Statistics (OES) Survey	12
Centers for Disease Control and Prevention (CDC)	
Behavioral Risk Factor Surveillance System (BRFSS)	14
National Ambulatory Medical Care Survey (NAMCS)	16
National Health and Nutrition Examination Survey (NHANES)	18
National Health Interview Survey (NHIS)	20
National Hospital Ambulatory Medical Care Survey (NHAMCS).....	21
National Hospital Care Survey (NHCS).....	22
National Hospital Discharge Survey (NHDS)	23
National Post-Acute and Long-Term Care Study (NPALS)	24
National Survey of Family Growth (NSFG).....	25
National Survey of Residential Care Facilities (NSRCF)	26
National Youth Tobacco Survey (NYTS)	27
Pregnancy Risk Assessment Monitoring System (PRAMS)	28
State and Local Area Integrated Telephone Survey (SLAITS)	29
Youth Risk Behavior Surveillance System (YRBSS)	30
Centers for Medicare and Medicaid Services (CMS)	
Health Outcomes Survey (HOS)	32
Medicaid Claims Data (National Medicaid Analytic eXtract [MAX] Data).....	33
Medicare Claims Data	35
Medicare Current Beneficiary Survey (MCBS).....	37
National Provider Identifier (NPI) File	38
Outcome and Assessment Information Set (OASIS)	39
Research Data Assistance Center (ResDAC)	40

Data Resource Center for Child and Adolescent Health (DRC)	
National Survey of Children with Special Health Care Needs (NS-CSHCN)	42
National Survey of Children's Health (NSCH)	43
Health Resources and Services Administration (HRSA)	
Area Health Resources Files (AHRF)	46
Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps).....	47
HRSA Data Warehouse	49
National Sample Survey of Nurse Practitioners (NSSNP).....	50
National Sample Survey of Registered Nurses (NSSRN).....	51
Uniform Data System (UDS).....	52
National Center for Education Statistics (NCES)	
Integrated Postsecondary Education Data System (IPEDS)	55
Substance Abuse and Mental Health Services Administration (SAMHSA)	
National Mental Health Services Survey (N-MHSS).....	57
National Survey of Substance Abuse Treatment Services (N-SSATS).....	58
National Survey on Drug Use and Health (NSDUH)	59
US Census Bureau	
American Community Survey (ACS)	61
Current Population Survey (CPS)	63
Economic Census	65
Survey of Business Owners and Self-Employed Persons (SBO)	66
Survey of Income and Program Participation (SIPP)	67
APPENDIX: Summary List of Federal Data Sources for Health Workforce Analysis	68



FEDERAL DATA SOURCES

INTRODUCTION

The US Government supports a wide range of surveys that collect data and information to inform public- and private-sector decision-making. This compendium provides a summary of 45 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 can 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 several nonfederal resources available that can also be used for health workforce analysis, but these resources are not summarized here. For a summary of all the resources reviewed in this compendium, see Table A-1 in the Appendix.

Agency for Healthcare
Research and Quality (AHRQ)

CONSUMER ASSESSMENT OF HEALTHCARE PROVIDERS AND SYSTEMS (CAHPS)

FEDERAL AGENCY: Agency for Healthcare Research and Quality (AHRQ) and US Department of Health and Human Services (DHHS)

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

DESCRIPTION: CAHPS is a series of surveys overseen by the AHRQ. They began in 1995 and include surveys administered by the Centers for Medicare and Medicaid Services (CMS) that follow CAHPS principles (<https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS>). CAHPS surveys ask patients to report on and assess their healthcare experiences with healthcare providers and plans, including hospitals, home healthcare 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 healthcare 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 healthcare 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 has a design tailored to its purpose.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The CAHPS survey results can be used by healthcare providers, healthcare organizations, insurers, regulators, educators, and other stakeholders to evaluate the quality and patient satisfaction of care provided, help target interventions to improve quality and satisfaction, and to track improvement. Although the surveys are designed for easy comparison across healthcare 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 varies depending on whether the survey was administered by 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 (<https://cahpsdatabase.ahrq.gov/DataResearchers.aspx>).

HEALTHCARE COST AND UTILIZATION PROJECT (HCUP)

FEDERAL AGENCY: Agency for Healthcare Research and Quality (AHRQ) and US Department of Health and Human Services (DHHS)

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

DESCRIPTION: HCUP is a group of healthcare 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 US, with data from approximately 7 million hospital stays (<https://www.hcup-us.ahrq.gov/nisoverview.jsp>). The KID, started in 1997, is a nationwide sample of pediatric inpatient discharges (<https://www.hcup-us.ahrq.gov/kidoverview.jsp>). The NEDS, started in 2006, contains more than 25 million records for emergency department visits at about 1,000 US community hospitals (<https://www.hcup-us.ahrq.gov/nedsoverview.jsp>).

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 (eg, 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 healthcare captured by the data. State-specific data files, where available, generally include records for the universe of the event captured in the data file (eg, 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 healthcare 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). 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.

MEDICAL EXPENDITURE PANEL SURVEY (MEPS)

FEDERAL AGENCY: Agency for Healthcare Research and Quality (AHRQ) and US Department of Health and Human Services (DHHS)

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

DESCRIPTION: MEPS was initiated in 1996 and comprises multiple large-scale surveys of individuals and families across the US and their medical providers (eg, doctors, hospitals, pharmacies, etc.) and employers. In essence, MEPS collects data on the various healthcare services that Americans use, the frequency of their use, and their associated costs, in addition to specifics on the health insurance plans offered by US 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 3 MEPS components are described below.

Household Component (HC)

The HC collects information from a sample of individuals and families across the US drawn from a nationally representative subsample of households participating in the National Health Interview Survey (NHIS) 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 12,000-15,000 families covering some 30,000-38,000 individuals (https://meps.ahrq.gov/mepsweb/survey_comp/hc_sample_size.jsp). The 2020 Full Year Consolidated File from the HC (the year available at the time of this writing) contains 26,847 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 3,400 state and local governments.

Medical Provider Component (MPC)

The MPC 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 healthcare 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 healthcare.

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 the 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/data_tools.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. 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). For employer-level analyses, researchers with approved projects can access restricted data files covering business establishments through the US Census Research Data Center Network (<https://www.census.gov/programs-surveys/ces/data/restricted-use-data.html>).

Bureau of Labor Statistics (BLS)

AMERICAN TIME USE SURVEY (ATUS)

FEDERAL AGENCY: Bureau of Labor Statistics (BLS), US Department of Labor (DOL)

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

DESCRIPTION: The ATUS, sponsored by the BLS and conducted by the US 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 3 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 (<https://www.bls.gov/tus/atususersguide.pdf>).

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The ATUS, along with additional health-related modules, can be used to describe how workers in different healthcare 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 healthcare professionals. The ATUS can also be used to understand health-related factors in the general population as indicators of potential demand for certain healthcare 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 are included in the ATUS-CPS data file.

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

CURRENT EMPLOYMENT STATISTICS (CES) SURVEY

FEDERAL AGENCY: Bureau of Labor Statistics (BLS), US Department of Labor (DOL)

WEBSITES: <https://www.bls.gov/ces>

DESCRIPTION: The 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 131,000 businesses and government agencies in the US, representing more than 697,000 individual worksites.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The CES program provides information on total employment, payroll costs, and hours worked in healthcare establishments as grouped by the NAICS code. The CES program provides information on the 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: The 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 in the CES data section (<https://www.bls.gov/ces/home.htm#data>). Summary tables and data on state and metro area employment, hours, and earnings are available in the SAE data section (<https://www.bls.gov/sae/home.htm#data>).

EMPLOYMENT PROJECTIONS (EP) PROGRAM

FEDERAL AGENCY: Bureau of Labor Statistics (BLS), US Department of Labor (DOL)

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

DESCRIPTION: The EP program at the BLS provides US labor market projections for the next 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 EP Methods Overview (<https://www.bls.gov/emp/documentation/projections-methods.htm>) or under the BLS Handbook of Methods (<https://www.bls.gov/opub/hom/emp/pdf/emp.pdf>).

The EP program uses survey data from several different sources to develop projections for the labor force, the macroeconomy, industry output and employment, and occupational employment. Data sources include the Current Population Survey (CPS), the Current Employment Statistics (CES) survey, the Occupational Employment and Wage Statistics (OEWS) survey, and the US 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.htm>). 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 800 occupations, as well as summary tables on selected subjects, are available at <https://www.bls.gov/emp/tables.htm>. Detailed data files for researchers are also available (<https://www.bls.gov/emp/tables.htm>). 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), US Department of Labor (DOL)

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

DESCRIPTION: The OES survey 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 (eg, 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 180,000-200,000 establishments are selected semiannually. OES estimates are based on 6 panels of establishment survey data and rely on data from approximately 1.2 million sampled establishments collected over a 3-year period. (For more information, visit <https://www.bls.gov/oes/mb3-methods.pdf>.)

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The OES has information on more than 60 health occupations. Data elements include the estimated total number of individuals employed in an occupation and their 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 hospitals, physician offices, home healthcare, outpatient care centers, and nursing care facilities (eg, <https://www.bls.gov/oes/current/oes291141.htm>).

GEOGRAPHICAL DETAIL: Data are reported for the nation, by state, and for metropolitan and nonmetropolitan 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 (<https://www.bls.gov/oes/tables.htm>).

Centers for Disease Control
and Prevention (CDC)

BEHAVIORAL RISK FACTOR SURVEILLANCE SYSTEM (BRFSS)

FEDERAL AGENCY: Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The BRFSS is a collaborative project of the CDC and US states and territories. The BRFSS is an ongoing data collection program designed to measure health risk behaviors, preventive health practices, and healthcare access in the American adult population (age 18 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.

Currently, the BRFSS conducts more than 400,000 interviews of adults annually.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: BRFSS is a potential source of data on healthcare utilization and access by the adult population and is 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 the 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 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.

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. These data files enable a primary analysis of the case-level data.

The BRFSS Web Enabled Analysis Tool (WEAT) allows users to create cross-tabulation analyses and logistic regression analyses from variables available in the WEAT variable lists (<https://nccd.cdc.gov/weat/#/analysis>). 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. These data files may be used for primary analysis of the case-level data.

NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The NAMCS is an annual national survey designed to provide objective, reliable information on the availability and utilization of ambulatory medical care services in the US. The survey was conducted annually from 1973-1981, 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 1-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 has also 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 nonphysician 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 was intended, in part, to enable estimates for a majority of the states in the US.

AVAILABILITY: Documentation on NAMCS data files and downloadable PUFs are available from 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/b1/datatype/Dt1224a.htm>).

NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The NHANES is a multifaceted study designed to evaluate the health and nutritional status of US adults and children through a combination of interviews and physical examinations. NHANES is a key program of the 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 US 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 2-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 healthcare services, such as those associated with mental health and chronic conditions (eg, 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 HEALTH INTERVIEW SURVEY (NHIS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cdc.gov/nchs/nhis>

DESCRIPTION: The NHIS was initiated in 1957 to collect national data for monitoring the health of the population in the 50 US states and the District of Columbia. The NHIS collects data on health status, access to healthcare, care needs, the quality of healthcare, and other socioeconomic information. The NHIS has been widely used by researchers and practitioners who explore public health issues and advise health policymakers.

The NHIS is a cross-sectional household survey that includes noninstitutionalized populations, such as homeless shelters and rooming houses, but does not include homeless and transient persons not residing in shelters, military personnel and people living in military bases, people in long-term care facilities, correctional facilities, or foreign countries. The responses are collected based on a face-to-face interview format.

The NHIS employs geographically clustered sampling techniques to ensure each month's sample is nationally representative. The sampling plan of the NHIS is updated every decennial census. The current 2016-2025 NHIS is based on a sampling plan that reflects on the results of the 2010 decennial census. The NHIS updated its questionnaire and structure in 2019 to improve measurement on a variety of health topics and reduce the burden on respondents. Currently, the NHIS collects information on almost 40,000 persons annually (30,000 sampled adults and 9,000 sampled children).

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NHIS data can be used to monitor and investigate population health and examine the effects of federal and state health programs. Data from the NHIS do not include specific health workforce information.

GEOGRAPHICAL DETAIL: The restricted use of the NHIS provides state, county, and census block codes. The sample is representative at the national and census-region level.

AVAILABILITY: The data sets without state identifiers are publicly available. The restricted data sets are available upon request but only accessible at NCHS research data centers (RDCs) or Federal Statistical RDCs (FDRDCs). Additionally, survey questionnaires, methodologies, and variable lists of the NHIS are available at <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>. Detailed variable lists for the restricted use files are available at <https://www.cdc.gov/rdc/b1datatype/Dt1225.htm>.

NATIONAL HOSPITAL AMBULATORY MEDICAL CARE SURVEY (NHAMCS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The NHAMCS collects data annually on the availability and use of ambulatory care services in hospital emergency and outpatient departments. Since 1992, NHAMCS has obtained data from a national sample of visits to the emergency departments (EDs), outpatient departments (OPDs), and ambulatory surgery locations (ASLs) of noninstitutional 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 350-400 hospitals have participated annually in the NHAMCS, yielding about 30,000-35,000 encounter (visit) records annually for ED visits; 30,000-35,000 records for OPD visits; and 15,000-20,000 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/datasets_documentation_related.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/b1datatype/Dt1224a.htm>).

NATIONAL HOSPITAL CARE SURVEY (NHCS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cdc.gov/nchs/nhcs/index.htm>

DESCRIPTION: The NHCS was designed to monitor the latest trends affecting hospitals and healthcare organizations, as well as factors associated with the healthcare utilization, the quality of healthcare, and disparities in healthcare services in the US. The NHCS collects data on patient care in hospital-based settings, such as inpatient (IPs), emergency departments (EDs), and outpatient departments (OPDs), and monitors national trends in substance use-related ED visits. The NHCS also includes information on diagnoses, services, discharge status, point of origin, start and end date of the hospital encounter, and patient identifiers. The patient identifiers allow linking with external datasets, such as outside data sources, including the National Death Index (NDI) and Centers for Medicare and Medicaid Services data.

Currently, the data is available for 4 years, from 2013-2016. The data collection for these years was based on a national probability sample that consisted of 581 hospitals. The sample of the NHCS includes hospitals with 6 or more beds staffed for inpatient care in the 50 states and the District of Columbia. However, the response rates were not high enough to produce weighted national estimates.

The NHCS was launched in 2011. For that year, the sample frame was based on the 2010 IMS Government Solutions hospital database. A sample of 1,000 hospitals was chosen from the original frame file of 6,622 hospitals. The sample was chosen using a stratified approach based on hospital bed size, hospital type, and being in urban/rural areas. These 1,000 hospitals were divided into 2 groups of 500, a base sample and a reserve sample. In 2013, 81 general acute hospitals with at least 500 staffed inpatient beds from the reserve sample were added to the base sample. From 2013-2016, 581 hospitals were eligible to participate in the survey.

At the time of this publication, preliminary data for the years 2020-2022 were available on a limited basis.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NHCS data can be used to track the trends of the factors affecting hospitals and healthcare delivery and the care received in hospitals. Data from the NHCS do not include specific health workforce information.

GEOGRAPHICAL DETAIL: The NHCS is suitable for national- and regional-level analyses. The data only provide 4 regional categories (Northeast, Midwest, South, and West), without state or county identifiers.

AVAILABILITY: The NHCS data are only available through the NCHS Research Data Center (<https://www.cdc.gov/rdc/index.htm>). Additionally, PDFs of the questionnaire, codebook, and methodology report from each year are available at <https://www.cdc.gov/rdc/b1datatype/dt1224h.htm>.

NATIONAL HOSPITAL DISCHARGE SURVEY (NHDS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: Conducted annually from 1965-2010, the NHDS was a national probability survey that collected data on the characteristics of inpatients discharged from nonfederal, noninstitutional, short-stay hospitals in the US. It employed 2 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 US 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 automated 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. 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/b1datatype/dt1224b.htm>).

NATIONAL POST-ACUTE AND LONG-TERM CARE STUDY (NPALS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cdc.gov/nchs/npals/index.htm>

DESCRIPTION: In January 2020, the National Study of Long-Term Care Providers (NSLTCP) was renamed the National Post-acute and Long-term Care Study (NPALS). The NPALS monitors trends in the supply, delivery, and utilization of paid, regulated long-term care services. It 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 2 major components, the Residential Care Community Questionnaire and the Adult Day Services Center Questionnaire (<https://www.cdc.gov/nchs/npals/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. The sampling frame 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 (CMS).

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NPALS 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: NPALS has conducted 5 waves of data collection since 2012. Reports on the first 4 waves are available at <https://www.cdc.gov/nchs/npals/reports.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/b1datatype/dt1224f.htm>. These data sets are not public and require the submission of a proposal to the RDC and 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.

NATIONAL SURVEY OF FAMILY GROWTH (NSFG)

FEDERAL AGENCY: Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cdc.gov/nchs/nsfg/index.htm>

DESCRIPTION: The NSFG was launched in 1973 to produce reliable evidence on predictors of and attitudes about various family issues, including pregnancy, live birth, marriage, divorce, cohabitation, family building, contraception, infertility, child adoption, caring, and the use of sexual and reproductive health services. The survey collects national data on pregnancy and births, marriage and cohabitation, infertility, use of contraception, family life, and general and reproductive health.

The NSFG changed to a continuous survey design starting in 2006 (ie, 2006-2010, 2011-2013, 2013-2015, 2015-2017, and 2017-2019), after it had been a periodically conducted survey design from 1973-2002 (ie, 1973, 1976, 1982, 1988, 1995, and 2002). For example, the 2017-2019 data, which are the most recent release, NSFG includes responses collected in 8 quarters over 2 years. The responses are collected based on in-person interviews. The overall response rate for the 2017-2019 datasets was 63.4%.

The NSFG is based on a multistage, probability-based, nationally representative sample of the household population ages 15-49. In 1973, the NSFG initially collected responses from a nationally representative sample of ever-married women ages 15-44 in the civilian, noninstitutionalized population of the US. The sample then expanded to interview women ages 15-44 regardless of marital experience in 1982, and men ages 15-44 in 2002, and women and men age 49 and older in 2015.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NSFG data can be used to study family growth/life, marriage, pregnancy, births, and reproductive health, and their predictors. Evidence from such research can inform public health agencies and officials from federal, state, and local governments to plan health-related social service programs. However, these data do not include specific information on the health workforce.

GEOGRAPHICAL DETAIL: The NSFG is suitable for national, state, and county analyses. The restricted-use version of the NSFG provides respondents' state or county of residence.

AVAILABILITY: The NSFG data sets without state- and county-identifiers are available for public use; data are available in SAS, SPSS, and STATA formats. Additionally, PDFs of the questionnaire, codebook, and methodology report from each year are available at https://www.cdc.gov/nchs/nsfg/nsfg_questionnaires.htm.

NATIONAL SURVEY OF RESIDENTIAL CARE FACILITIES (NSRCF)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The 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 4 or more registered beds, provide room and board with at least 2 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 2 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 US, 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 https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/.

Before downloading the files, data users can review the information at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileDataReadMe.pdf and https://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), US Department of Health and Human Services (DHHS)

WEBSITE: https://www.cdc.gov/tobacco/data_statistics/surveys/nyts

DESCRIPTION: The 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 6-12. The NYTS also serves as a reference point for meeting certain Healthy People 2030 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 (2021), a stratified, 3-stage cluster sample design was used to yield a nationally representative sample of US 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 can also be used to estimate the need for services such as smoking cessation 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 and Excel 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.

PREGNANCY RISK ASSESSMENT MONITORING SYSTEM (PRAMS)

FEDERAL AGENCY: Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The 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, 81% of all US 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 monthly by states 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 policy-makers 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 healthcare 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 US 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 at the CDC's All Chronic Disease and Health Promotion Data & Indicators (<https://chronicdata.cdc.gov/browse>). 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/prams/researchers.htm>. Proposal reviews are conducted once per month.

STATE AND LOCAL AREA INTEGRATED TELEPHONE SURVEY (SLAITS)

FEDERAL AGENCY: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cdc.gov/nchs/slaits/index.htm>

DESCRIPTION: The SLAITS, which was initiated in 1997, includes various modules on health, child welfare, and care needs. The SLAITS data collection is funded by sponsorship of these specific modules for government agencies and nonprofit organizations. For example, the National Survey of Children with Special Health Care Needs, which is a module of the SLAITS, was funded by the Maternal and Child Health Bureau of the Health Resources and Services Administration to collect data to identify prevalent and special healthcare needs among children. Other survey modules include health, child well-being and welfare, childhood health, asthma, influenza vaccination, adoptive parents, children with special healthcare needs, children in nonparental care, and ADHD and Tourette syndrome.

The SLAITS modules vary in topics and populations studied. The first module on health gathered data from Iowa and Washington in 1997. The second module on child well-being and welfare collected responses in Texas and Minnesota. Other modules have included national samples and state samples. The most recent module is the National Survey of the Diagnosis and Treatment of ADHD and Tourette Syndrome, which was collected in 2014.

The SLAITS modules use a telephone sampling frame, including a cellphone sample, of the CDC's National Immunization Survey (NIS). The NIS has used random-digit-dial sample and computer-assisted telephone interview approaches to collect responses in 50 states and the District of Columbia.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The SLAITS modules can be used to investigate various health issues, including general health, child well-being, and child welfare. Data from the SLAITS do not include specific health workforce information.

GEOGRAPHICAL DETAIL: The SLAITS modules are suitable for national- and state-level, cross-sectional analysis. State governments can use these data for comparing various health issues across the state. However, most of the modules are not regularly collected.

AVAILABILITY: The SLAITS data sets are available for public use in SAS format. Additionally, PDFs of the questionnaire, codebook, and methodology report from each year are available at <https://www.cdc.gov/nchs/slaits/index.htm>.

YOUTH RISK BEHAVIOR SURVEILLANCE SYSTEM (YRBSS)

FEDERAL AGENCY: Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The 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 grades 9-12 students via national, state, territorial, tribal government, and local school-based surveys administered biennially. Between 1991-2019, the YRBSS used more than 2,100 surveys to collect data from more than 4.9 million high school students.

For each state, territorial, tribal, and large urban school district, the YRBSS utilizes a 2-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 3-part, cluster sample design. National YRBSS data are representative of all public and private high school students in all 50 states, the District of Columbia, and some US territories. 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 US territories, for large urban school districts, and for tribal governments. The availability of a geographical level below the national level 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 ASCII and Microsoft Access. SAS and SPSS programs are also available.

Centers for Medicare and Medicaid
Services (CMS)

HEALTH OUTCOME SURVEY (HOS)

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The HOS is designed to collect valid, reliable, and meaningful health-status data from beneficiaries of 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. The HOS also collects information on 3 Healthcare Effectiveness Data and Information Set (HEDIS) effectiveness of care parameters: Management of Urinary Incontinence in Older Adults, Physical Activity in Older Adults, and Fall Risk Management.

The HOS is administered each year to a random sample of Medicare beneficiaries drawn from participating MA plans. Follow-up surveys are conducted after 2 years. All managed care organizations with Medicare contracts are mandated to participate.

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 demand for healthcare 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 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 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>

MEDICAID CLAIMS DATA (MEDICAID ANALYTIC EXTRACT [MAX] DATA)

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: Medicaid claims-level data for the years 1999-2015 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.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MSIS>). The MAX is a set of person-level data files on Medicaid eligibility, service utilization, and payments. 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 (eg, 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 and 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 in 2009 data, a MAX file on provider characteristics (MAXPC) was also 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 healthcare, case management, personal care, nonemergency transportation, and many other services are excluded from the NPI requirement. MAXPC, however, does include legacy identification numbers in addition to the NPI. In addition to information about the provider, MAXPC contains information on the number of different claims (eg, 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: MAX data can be used to assess provider supply and healthcare utilization for the Medicaid population. Provider information in claims data may enable the 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 (ie, personal summary data set), claims data can also be used to assess rates of utilization for different services across individual demographic variables in the data.

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 personally identifiable information, availability is restricted by the Privacy Rule. MAX data files are available for approved research only and a Data Use Agreement with CMS is required. For help in preparing a data request, contact the Research Data Assistance Center (<https://resdac.org/>).

MAXPC is downloadable directly from the CMS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXPC.html>.

MEDICARE CLAIMS DATA

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US 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 data files contain information collected by Medicare to pay for healthcare 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 healthcare provider and facility data. Each Medicare claims data 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.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/bsapufs>.

In addition to the Medicare 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 Medicare claims data, enables analysis of the rate of Medicare utilization per beneficiary and related analyses. (For additional information, see <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MBSF-LDS>.)

Medicare claims data 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 healthcare 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 (ie, 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> 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://resdac.org/research-identifiable-files-rif-requests>.

LDS files also are available in 2 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>.

MEDICARE CURRENT BENEFICIARY SURVEY (MCBS)

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The 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 (National Opinion Research Center) 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 50 US states, 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, healthcare utilization, and costs for older adults (ie, 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 older adults and the volume of health utilization across different providers. The detailed data in the MCBS allow for estimating or predicting how demand for healthcare—and, by extension, for healthcare providers—among the older adult 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 healthcare and the health workforce.

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

AVAILABILITY: A PUF and accompanying documentation for the 2019 MCBS are 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 healthcare, 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), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The Health Insurance Portability and Accountability Act (HIPAA) requires HIPAA-covered healthcare providers to obtain and use an NPI, which is a unique identification number. HIPAA-covered healthcare 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 (eg, physicians, nurses, dentists, chiropractors, physical therapists, and pharmacists) and organizations (eg, 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 healthcare 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. 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/NationalProviderStand/DataDissemination.html>.

HOME HEALTH OUTCOME AND ASSESSMENT INFORMATION SET (OASIS)

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Home-Health-Quality-Reporting-Requirements>

DESCRIPTION: The OASIS collects data on the utilization of the Medicare Home Health (HH) benefit, socio-demographic characteristics of HH patients, and patient outcomes of home health agencies (HHAs). The variables include beneficiary ID; provider number; claim date; primary diagnosis; costs for services; and types of visits, services, and therapies. The OASIS is an establishment survey required by the law. The OASIS assessments for all patients ages 18 and older who receive skilled HH services are mandatory for Medicare or Medicaid-certified HHAs.

The OASIS data are updated twice a year. The first update is published in early summer and is related to the yearly notices of proposed rulemaking. The second update, which related to the final rule, is posted in the autumn. The data can be linked to other external sources, such as the Surveillance, Epidemiology and End Results (SEER) Medicare data and other CMS data on healthcare providers.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The OASIS data can be used to estimate the frequency of HH care use, patient characteristics, and patient outcomes. This information can help patients and customers compare the quality of HHAs. Data from the OASIS do not include specific health workforce information.

GEOGRAPHICAL DETAIL: The OASIS is suitable for national analysis; the data does not provide state or county identifiers. However, researchers can conduct subnational analysis by linking the data with other provider data administered by the CMS.

AVAILABILITY: The OASIS data sets are not available for public use. Data can be obtained by request from the Research Data Assistance Center (ResDAC) and are provided in a comma separated variable file with a SAS read-in program. The cost for obtaining the 2020 release, which is the most recent year, is \$1,200. PDFs of the questionnaire, codebook, and methodology report from each year are available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/OASIS-Data-Sets>.

RESEARCH DATA ASSISTANCE CENTER (RESDAC)

FEDERAL AGENCY: Centers for Medicare and Medicaid Services (CMS), US Department of Health and Human Services (DHHS)

WEBSITE: <https://resdac.org>

DESCRIPTION: The ResDAC is a data repository administered by the CMS. Data from the ResDAC contain information on the Medicare and Medicaid programs, including enrollments, payments, and quality of healthcare providers. Currently, the ResDAC provides 68 different data sources. The data coverages range from 1999-2020.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: Data from the ResDAC can be used to monitor Medicaid and Medicare enrollments, payments, and provider quality and improve Medicaid and Medicare programs. Several data sources from the ResDAC include specific health workforce information, including the skilled nursing facility data and the home health agency data.

GEOGRAPHICAL DETAIL: The ResDAC provides data sources that include subnational identifiers, such as states and counties.

AVAILABILITY: The ResDAC provides 3 types of data: public use files (PUF), limited data sets (LDS), and research identifiable files (RIF). PUF files typically contain aggregated data and can be obtained immediately or within 2 weeks. However, other types, which include individual level data, whether de-identified or not, are only available upon request, and may incur fees and take weeks or months to process.

Data Resource Center for Child and Adolescent Health (DRC)

NATIONAL SURVEY OF CHILDREN WITH SPECIAL HEALTH CARE NEEDS (NS-CSHCN)

ORGANIZATION: Data Resource Center for Child and Adolescent Health (DRC)

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

DESCRIPTION: The NS-CSHCN was designed to examine the health and functional status of children with special healthcare needs in the US. The survey was conducted 3 times between 2001-2010, and has been integrated into the National Survey of Children's Health (NSCH) as of 2016. The NS-CSHCN assessed the incidence and impact of special healthcare needs among children younger than age 18, 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 each survey year. All versions of the survey were weighted to be representative of the US population of noninstitutionalized children younger than age 18. 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. See <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 younger than age 18 who have special healthcare needs. The survey also contains information on the use of healthcare services, including whether, where, and how often the child goes for healthcare visits; accessibility of healthcare services; experience with the child's healthcare providers and care coordination; transition plans from pediatric to adult care; health insurance coverage and cost of the child's healthcare; and demographic, educational, and socioeconomic data. These data can be used in health workforce planning efforts to better understand the healthcare needs and utilization patterns of children with special healthcare needs and estimate the need for appropriate pediatric healthcare 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 include all variables from the NSCH public-use data files and can be requested at <http://www.childhealthdata.org/help/dataset>. Available file formats include SAS, SPSS, and STATA.

NATIONAL SURVEY OF CHILDREN'S HEALTH (NSCH)

ORGANIZATION: Data Resource Center for Child and Adolescent Health (DRC)

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

DESCRIPTION: The NSCH provides detailed data on children in the US younger than age 18. Conducted 3 times between 2003-2012, the survey covers numerous interconnected aspects of children's lives, including their physical and mental health; access to healthcare; and family, neighborhood, educational, and social environments. The survey has been conducted by the US Census Bureau since 2016. Notably, the National Survey of Children with Special Health Care Needs (see previous 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 US. 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 US 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 healthcare 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 younger than age 18, both in the US 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 younger than age 18. The NSCH also contains information on the use of healthcare services, including whether, where, and how often the child goes for healthcare visits; accessibility of healthcare services; experiences with the child's healthcare providers and care coordination; health insurance coverage and cost of the child's healthcare; and demographic, educational, and socioeconomic data. These data can be used in health workforce planning efforts to better understand children's healthcare needs and utilization patterns and estimate the need for pediatric healthcare 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 variable indicating 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 DRC. 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.

Health Resources and Services Administration (HRSA)

AREA HEALTH RESOURCES FILES (AHRF)

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

WEBSITE: <https://data.hrsa.gov/topics/health-workforce/ahrf>

DESCRIPTION: The 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. The data in the AHRF are compiled from more than 50 sources, including the American Medical Association, the American Hospital Association, the American Dental Association, the US Census Bureau, the Centers for Medicare and Medicaid Services, and the National Center for Health Statistics, among others. The AHRF is released annually.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The AHRF contains data on a range of healthcare providers and can be used 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 (eg, Federal Information Processing Standard d [FIPS] 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; and land area.

AVAILABILITY: AHRF data can be downloaded at no cost. Information and a download link can be found at <https://data.hrsa.gov/topics/health-workforce/ahrf>. The direct download site is <https://data.hrsa.gov/data/download?data=AHRF#AHRF>.

HEALTH PROFESSIONAL SHORTAGE AREAS (HPSAS) AND MEDICALLY UNDERSERVED AREAS/POPULATIONS (MUA/PS)

FEDERAL AGENCY: Health Resources and Services Administration (HRSA), US 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 HPSA or a 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 MUA/Ps are groups of people for whom healthcare 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 older adults, 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: HPSA and MUA/P designations are used to determine eligibility 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 HPSA and MUA/P programs are designed to help communities recruit and retain primary care medical providers (eg, physicians, nurse practitioners, physician assistants), nurses, dental providers, and mental and behavioral healthcare providers (eg, 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).

HPSA designations are used to recruit healthcare 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: HPSA and MUA/P data are available via the HRSA Data Warehouse, which allows users to interact with data in charts, tables, reports, and maps (<https://data.hrsa.gov/data/about>).

HRSA DATA WAREHOUSE (HDW)

FEDERAL AGENCY: Health Resources and Services Administration (HRSA), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The HDW provides maps, data, reports, and dashboards to the public. The data integrate with external sources, such as the US Census Bureau, providing information about HRSA's grants, loan and scholarship programs, health centers and other public health programs and services. Several data sets are available for download, including 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).

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: These data can be used for forecasting provider supply and demand, planning educational programs, developing regulatory policies, and evaluating the impact of policy decisions. The available data in the HRSA data warehouse can help planners and policy makers to prepare and build a health workforce that may 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, and reports. The HDW manages data from more than 30 HRSA and other federal government sources at <https://datawarehouse.hrsa.gov/default.aspx>.

NATIONAL SAMPLE SURVEY OF NURSE PRACTITIONERS (NSSNP)

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

WEBSITE: <https://bhw.hrsa.gov/data-research/access-data-tools/national-sample-survey-registered-nurse-practitioners>

DESCRIPTION: The NSSNP was conducted in 2012 to better understand the supply, distribution, and role of nurse practitioners (NPs) in the healthcare system. NPs have become increasingly relied upon in the US to provide high quality primary care, particularly as demand for services has risen. The NSSNP was integrated into the National Sample Survey of Registered Nurses (NSSRN) in 2018.

For the 2012 survey, HRSA consulted state licensing boards to obtain lists of all actively licensed NPs in the US. 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. 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://data.hrsa.gov/topics/health-workforce/nursing-workforce-survey-data?tab=NursePractitioners>. 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.

NATIONAL SAMPLE SURVEY OF REGISTERED NURSES (NSSRN)

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

WEBSITE: <https://bhwh.hrsa.gov/data-research/access-data-tools/national-sample-survey-registered-nurses>

DESCRIPTION: First conducted in 1977, HRSA's Bureau of Health Workforce conducted the NSSRN approximately every 4 years through 2008. The most recent edition of the survey was in 2018. The data from the NSSRN provide the opportunity to assess trends in and make projections on the nursing workforce. The survey asks registered nurse (RN) 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 RNs 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 RN's record, with adjustments for those RNs with multiple state licenses. Although some RNs 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.

The US Census Bureau was responsible for administering the 2018 NSSRN. The 2018 NSSRN also incorporated the previous National Sample Survey of Nurse Practitioners (NSSNP) into its design and, as a result, oversampled NPs in each state to ensure an adequate representation. In addition, the Census Person Identification Validation System (PVS) was used to improve the 2018 sampling frame by making demographic data (eg, race/ethnicity) available for sample selection and sample weight adjustment.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NSSRN provides information on the supply and characteristics of RNs in the US. 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 RNs.

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

AVAILABILITY: For each survey year, HRSA has prepared public-use data files (PUFs) in the ASCII file format. SAS, SPSS, and STATA file formats are also available for some years. PUFs and documentation for any of the NSSRN surveys may be downloaded at <https://data.hrsa.gov/topics/health-workforce/nursing-workforce-survey-data>.

UNIFORM DATA SYSTEM (UDS)

FEDERAL AGENCY: Health Resources and Services Administration (HRSA), US Department of Health and Human Services (DHHS)

WEBSITE: <https://bphc.hrsa.gov/datareporting/reporting/index.html>

DESCRIPTION: The UDS is a data set that is based on an annual program requirement for the HRSA Health Center Program awardees, as established in Public Health Service Act Section 330. Health centers are required to provide annual reports to the UDS on patient profiles, services delivered, clinical procedures and outcomes, staffing, patient use of services, and expenses and revenues. These health centers must adhere to Health and Human Services (HHS) reporting standards and must have a system in place to produce and manage data pertaining to the HRSA-approved scope of work.

The UDS includes 11 tables and 3 appendices that health centers must complete. They are:

- ZIP Codes: Patients by ZIP code and by primary third-party medical insurance source
- Table 3A: Patients by age and by sex assigned at birth
- Table 3B: Patients by race, Hispanic or Latino/a ethnicity, language barriers, sexual orientation, and gender identity
- Table 4: Patients by income and primary third-party medical insurance source, the number of special population patients receiving services, and managed-care enrollment
- Table 5: The annualized full-time equivalent (FTE) of program staff by position category, visits by provider type, patients by service type, and mental health and substance use disorder selected services detail
- Table 6A: Visits and patients for selected medical, mental health, substance use disorder, vision, and dental diagnoses and services
- Table 6B: Clinical quality-of-care measures
- Table 7: Health outcome measures by race and ethnicity
- Table 8A: Direct and indirect expenses by service categories
- Table 9D: Full charges, collections, and adjustments by payer type; sliding fee discounts; and bad debt write-offs for patients
- Table 9E: Other, nonpatient-service-related income
- Appendix D: Health Information Technology (HIT) Capabilities Form: HIT capabilities, including the use of electronic health record (EHR) information

- Appendix E: Other Data Elements Form: Medication-assisted treatment (MAT), telehealth, and outreach and enrollment assists
- Appendix F: Workforce Form: Health center workforce training and provider and staff satisfaction surveys

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The UDS data can be used to investigate various public health issues, such as healthcare service delivery and usage. These data can also be used to study issues related to access to care, health disparities, quality of care, and healthcare costs, among other topics.

GEOGRAPHICAL DETAIL: The UDS includes identifiers for HHS region, state, county, and congressional district.

AVAILABILITY: UDS data sets are available for public use; data are available in Microsoft Excel and comma-separated values formats. Additionally, PDFs of the codebook and methodology report for each year are available at <https://data.hrsa.gov/tools/data-reporting>. Additionally, HRSA periodically provides analytic reports based on the UDS data through the website of the Bureau of Primary Health Care (BPHC): <https://bphc.hrsa.gov/datareporting/index.html>.

**National Center for Education
Statistics (NCES)**

INTEGRATED POSTSECONDARY EDUCATION DATA SYSTEM (IPEDS)

FEDERAL AGENCY: National Center for Education Statistics (NCES), US Department of Education

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

DESCRIPTION: The IPEDS comprises a series of interrelated surveys administered annually by the NCES, part of the US Department of Education. IPEDS collects institutional-level data from postsecondary institutions in the US (50 states and the District of Columbia) and other US 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 from postsecondary institutions 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 healthcare 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 healthcare 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, bachelor’s, master’s, and doctoral) 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 (ie, 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>.

**Substance Abuse and Mental Health
Services Administration (SAMHSA)**

NATIONAL MENTAL HEALTH SERVICES SURVEY (N-MHSS)

FEDERAL AGENCY: Substance Abuse and Mental Health Services Administration (SAMHSA), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.samhsa.gov/data/data-we-collect/n-mhss-national-mental-health-services-survey>

DESCRIPTION: The N-MHSS is an annual survey of all publicly and privately operated specialty mental healthcare facilities in the US 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 (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.

Reports describing the methodology and findings from the survey are available at https://www.samhsa.gov/data/data-we-collect/n-mhss-national-mental-health-services-survey#report_type_1.

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: The data includes state identifiers.

AVAILABILITY: Public use files in various formats can be downloaded directly from SAMHSA's N-MHSS website at <https://www.datafiles.samhsa.gov/dataset/national-mental-health-services-survey-2020-n-mhss-2020-ds0001>.

NATIONAL SURVEY OF SUBSTANCE ABUSE TREATMENT SERVICES (N-SSATS)

FEDERAL AGENCY: Substance Abuse and Mental Health Services Administration (SAMHSA), US Department of Health and Human Services (DHHS)

WEBSITE: <https://www.samhsa.gov/data/data-we-collect/n-ssats-national-survey-substance-abuse-treatment-services>

DESCRIPTION: The N-SSATS is an annual survey of all public and private substance abuse treatment facilities in the US. The survey collects information about the facilities, services offered, and utilization of services. The information is used by SAMHSA, state and local governments, and other stakeholders, for program administration and policy analysis.

Reports describing the methodology and findings from the survey are available at https://www.samhsa.gov/data/data-we-collect/n-mhss-national-mental-health-services-survey#report_type_1.

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 addiction treatment programs), and client counts (eg, 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: The data can be analyzed at the national and state levels.

AVAILABILITY: Public use files in various formats can be downloaded directly from SAMHSA's N-SSATS website <https://www.datafiles.samhsa.gov/dataset/national-survey-substance-abuse-treatment-services-2020-n-ssats-2020-ds0001>.

NATIONAL SURVEY ON DRUG USE AND HEALTH (NSDUH)

FEDERAL AGENCY: Substance Abuse and Mental Health Services Administration (SAMHSA), US Department of Health and Human Services (DHHS)

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

DESCRIPTION: The NSDUH, conducted by SAMHSA, collects data on the use of illegal drugs, alcohol, and tobacco by noninstitutionalized US residents ages 12 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/data-we-collect/nsduh-national-survey-drug-use-and-health#report_type_2.

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 for direct download on SAMHSA's NSDUH website <https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2020-nsduh-2020-ds0001>.

US Census Bureau

AMERICAN COMMUNITY SURVEY (ACS)

FEDERAL AGENCY: US Census Bureau, US Department of Commerce

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

DESCRIPTION: The ACS is an ongoing statistical survey that samples a small percentage of the population in the US 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 4 types of ACS data files: 1-year files, 1-year supplemental estimates, 3-year files, and 5-year files. The 1-year files (and supplemental estimates) contain data for a single year, the 3-year files combine data from 3 years, and the 5-year files combine 5 years of ACS data. The 3-year files were discontinued after the 2011-2013 release.

The ACS annually collects data from about 2 million households. The sample is based on the US Census Bureau's official inventory of known living quarters, called 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 1 in 38 households receives an invitation to participate in the ACS each year (<https://www.maine.gov/dafs/economist/sites/maine.gov.dafs.economist/files/inline-files/ACS%20Information%20Guide-compressed.pdf>).

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 US Census Bureau classification scheme (based upon SOC) (eg, see <https://www2.census.gov/programs-surveys/cps/methodology/Occupation%20Codes.pdf>). 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 healthcare needs underlying demand for healthcare providers.

GEOGRAPHICAL DETAIL: The geographical detail in estimates published in summary form by the US Census Bureau varies based on whether estimates are from 1-year, 1-year supplement, 3-year, or 5-year files. The

reliability of estimates for geographical units with the smallest population numbers increases as the number of years combined in a file increase. The 1-year files are available for geographic areas with at least 65,000 persons and the 1-year supplemental estimates are available for geographic areas with at least 20,000 persons, albeit the data are not as precise. The 3-year files are also available for geographic areas with at least 20,000 persons. The 5-year files have data available for all geographic areas and are the largest and most reliable samples (<https://www.census.gov/programs-surveys/acs/guidance/estimates.html>).

In addition, the Public-Use Microdata Samples (PUMS) contain individual household- and person-level data. The smallest geographical units of analysis are areas with a population of at least 100,000 persons (<https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html>).

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

ACS data are available in predefined tabulations of characteristics. The basic unit of analysis is a specific geographic entity (eg, a state or county) for which estimates of persons, families, households, or other characteristics are available. A user can select specific predefined tabulations through the US Census Bureau's main data dissemination tool (<https://data.census.gov/cedsci/>). More advanced users can directly download ACS and PUMS data files on the US Census Bureau's file transfer protocol (FTP) website (<https://www.census.gov/programs-surveys/acs/data/data-via-ftp.html>).

Information on accessing ACS data 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: US Census Bureau, US Department of Commerce; Bureau of Labor Statistics (BLS), US Department of Labor (DOL)

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

DESCRIPTION: The CPS is sponsored jointly by the US 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 US Census Bureau using a monthly probability sample of occupied housing units. The actual fieldwork 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 4 consecutive months; after a subsequent 8-month dormant period, households are surveyed again for another 4 consecutive months before finally leaving the sample. A personal-visit interview is required for all first-month households in the sample. 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 household's 8-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 ages 16 or older, resulting in about 150,000 person-level records in each monthly file.

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 data collection. 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 instance, 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 (eg, metropolitan or nonmetropolitan).

AVAILABILITY: Basic monthly CPS data files and the annual March and Annual Social and Economic (ASEC) Supplement data files, along with documentation, can be downloaded from <https://www.census.gov/programs-surveys/cps/data/datasets.html>. Summary data tables from the latest ASEC Supplement are available at <https://www.census.gov/programs-surveys/cps/data/tables.html>. Summary data files and tables on labor force statistics are also available from the BLS at <https://www.bls.gov/cps/tables.htm>.

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

ECONOMIC CENSUS (EC)

FEDERAL AGENCY: US Census Bureau, US Department of Commerce

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

DESCRIPTION: The EC provides comprehensive data on the American economy every 5 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 (eg, gross domestic product, production and price indices, 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 US, across nearly every industry in the private, nonfarm economy, and in every geographic area, including Puerto Rico and other US territories. This sample is chosen using a stratified sampling procedure with strata based on industry and geography. Basic data for nonselected 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 the 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 multiunit firm.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The EC contains information that describes the characteristics of various healthcare establishments by industry (eg, health care and social assistance), industry subsector (eg, ambulatory health care services, hospitals, nursing and residential care facilities, social assistance), and industry level (eg, offices of dentists), and by establishment size. Data include variables such as the number of establishments, receipts and revenue, annual payroll, and the number of paid employees. While there are limitations to using EC data to characterize the supply of healthcare professionals, the EC can provide a helpful context in broader research and planning initiatives.

GEOGRAPHICAL DETAIL: The EC gives statistics at 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 available to avoid revealing information about individual firms.

AVAILABILITY: EC statistics are available at <https://www.census.gov/programs-surveys/economic-census/data/datasets.html>. Additionally, EC tables are downloadable from the US 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: US Census Bureau, US Department of Commerce

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

DESCRIPTION: The SBO provides comprehensive information on selected economic and demographic characteristics for businesses and business owners. Conducted every 5 years since 1972, the SBO provides data on gender, ethnicity, race, and veteran status of business owners. Title 13 of the US 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 on an establishment basis. A company or firm is 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 healthcare businesses by industry (eg, health care and social assistance), industry subsector (eg, ambulatory health care services, hospitals, nursing and residential care facilities, social assistance), and industry level (eg, offices of dentists). Data include estimates of business ownership by gender, ethnicity, race, and veteran status; the number of firms with and without paid employees; the number of paid employees; and annual payroll. While there are limitations to using SBO data to characterize the supply of healthcare professionals, the SBO can provide a 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 US Census Bureau's website at <https://www.census.gov/programs-surveys/sbo/data.html>. The US Census Bureau's online, self-service data access tool (<https://data.census.gov/cedsci/>) allows selective retrieval and downloading of the SBO data.

SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)

FEDERAL AGENCY: US Census Bureau, US Department of Commerce

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

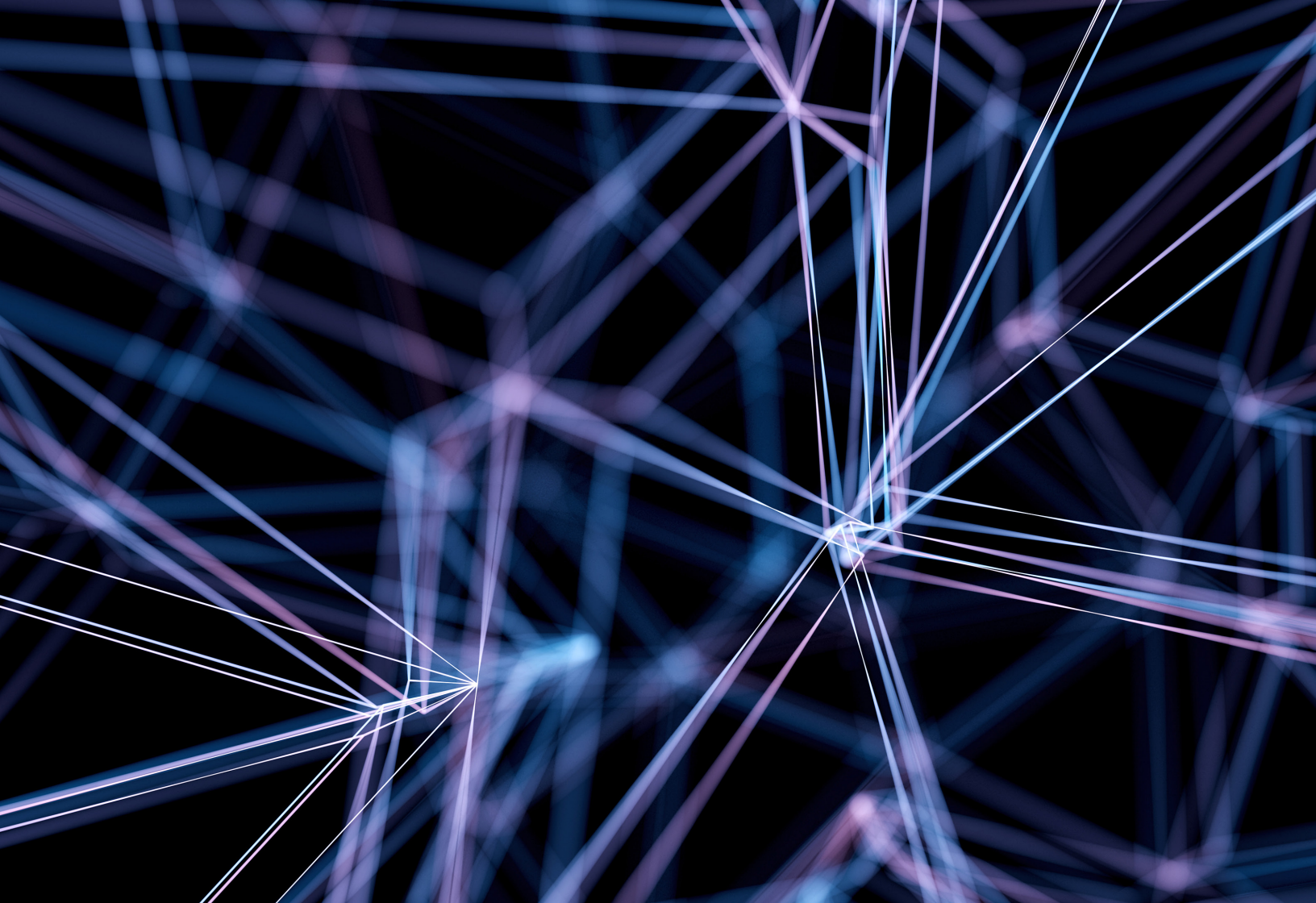
DESCRIPTION: The SIPP collects detailed data on the financial situations of households and individuals, including employment status and income, as well as more comprehensive metrics of well-being, such as family dynamics, educational attainment, housing, and utility expenses, assets, health insurance, 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 4-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. The survey design, sampling, and weighting are available in the SIPP methodology section <https://www.census.gov/programs-surveys/sipp/methodology.html>.

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

AVAILABILITY: SIPP data and documentation are public information, as released by the US Census Bureau, and downloadable SAS, STATA, delimited text files are available at <https://www.census.gov/programs-surveys/sipp/data/datasets.html>.



APPENDIX

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
Consumer Assessment of Healthcare Providers and Systems (CAHPS)	AHRQ, DHHS	https://www.ahrq.gov/cahps	Patient survey	Individual	Healthcare quality and satisfaction
Healthcare Cost and Utilization Project (HCUP)	AHRQ, DHHS	https://www.ahrq.gov/research/data/hcup	Health claims data	Healthcare encounter	Healthcare demand
Medical Expenditure Panel Survey (MEPS)	AHRQ, DHHS	http://meps.ahrq.gov/mepsweb	Household survey with component on sampled individual's health utilization data	Individual	Healthcare demand
American Time Use Survey (ATUS)	BLS, DOL	https://www.bls.gov/tus	Household survey	Individual	Time spent on work and nonwork activities
Current Employment Statistics (CES) Survey	BLS, DOL	https://www.bls.gov/ces	Payroll/establishment survey (<i>businesses and government agencies</i>)	Total employment by industry	Provider supply
Employment Projections (EP) Program	BLS, DOL	http://www.bls.gov/emp	Projection model	Occupation	Provider supply
Occupational Employment Statistics (OES) Survey	BLS, DOL	https://www.bls.gov/oes	Establishment mail survey	Occupation	Provider supply
Behavioral Risk Factor Surveillance System (BRFSS)	CDC, DHHS	https://www.cdc.gov/BRFSS	Telephone survey	Individual	Healthcare demand and access

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis (cont.)

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
National Ambulatory Medical Care Survey (NAMCS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/ahcd	Office-based physician survey	Healthcare encounter and physician	Provider supply; healthcare demand
National Health and Nutrition Examination Survey (NHANES)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nhanes	Survey	Individual	Healthcare demand
National Health Interview Survey (NHIS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nhis/index.htm	Household survey	Households & individuals; special populations	Healthcare access
National Hospital Ambulatory Medical Care Survey (NHAMCS)	NCHS, CDC, DHHS	http://www.cdc.gov/nchs/ahcd	Establishment survey	Healthcare encounter	Healthcare demand
National Hospital Care Survey (NHCS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nhcs/index.htm	Establishment survey	Healthcare encounter	Healthcare utilization
National Hospital Discharge Survey (NHDS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nhds	Establishment survey	Healthcare encounter	Healthcare demand
National Post-acute and Long-Term Care Study (NPALS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/npals/index.htm	Establishment survey	Establishment	Provider staffing; healthcare demand
National Survey of Family Growth (NSFG)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nsfg/index.htm	Household survey	Individuals	Healthcare planning
National Survey of Residential Care Facilities (NSRCF)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/nsrcf	Survey	Individual healthcare provider	Provider staffing; healthcare demand

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis (cont.)

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
National Youth Tobacco Survey (NYTS)	CDC, DHHS	https://www.cdc.gov/tobacco/data_statistics/surveys/nyts	Establishment survey	Establishment	Healthcare demand
Pregnancy Risk Assessment Monitoring System (PRAMS)	CDC, DHHS	https://www.cdc.gov/prams	Survey of women after childbirth	Individual	Healthcare demand
State and Local Area Integrated Telephone Survey (SLAITS)	NCHS, CDC, DHHS	https://www.cdc.gov/nchs/slaitis/index.htm	Telephone survey	Individual	Healthcare demand and access
Youth Risk Behavior Surveillance System (YRBSS)	CDC, DHHS	https://www.cdc.gov/healthyyouth/data/yrbs	Youth survey	Individual	Healthcare demand
Health Outcomes Survey (HOS)	CMS, DHHS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS	Medicare beneficiary survey	Individual	Healthcare demand
Medicaid Claims Data (Medicaid Analytic eXtract [MAX] Data)	CMS, DHHS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-GeneralInformation	Medicaid claims data	Individual	Provider supply; healthcare demand
Medicare Claims Data	CMS, DHHS	http://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html	Health claims data	Healthcare encounter	Provider supply; healthcare demand
Medicare Current Beneficiary Survey (MCBS)	CMS, DHHS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS	Individual survey supplemented with claims data	Individual	Healthcare demand and access
National Provider Identifier (NPI) File	CMS, DHHS	https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProviderIdentifier-Stand	Enrollment in program by healthcare providers	Individual health provider	Provider supply

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis (cont.)

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
Outcome and Assessment Information Set (OASIS)	CMS, DHHS	https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/homehealth-qualityinits	Establishment survey	Individuals	Healthcare demand
Research Data Assistance Center (ResDAC)	CMS, DHHS	https://resdac.org/	Compilation of multiple data sources to populate	Multiple units of analysis	Provider supply; healthcare demand and access
National Survey of Children with Special Health Care Needs (NS-CSHCN)	DRC	http://www.childhealthdata.org/learn/NS-CSHCN	Household survey	Individual children	Healthcare demand
National Survey of Children's Health (NSCH)	DRC	http://www.childhealthdata.org/learn/NSCH	Household survey	Individual children	Healthcare demand
Area Health Resources Files (AHRF)	NCHWA, HRSA, DHHS	https://data.hrsa.gov/topics/health-workforce/ahrf	Compilation of multiple data sources	County	Provider supply
Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas/Populations (MUA/Ps)	HRSA, DHHS	https://bhwh.hrsa.gov/shortage-designation	Health workforce shortage designations	Geographic areas; special populations	Healthcare demand
HRSA Data Warehouse	HRSA, DHHS	https://datawarehouse.hrsa.gov/default.aspx	Compilation of multiple data sources	Multiple units of analysis	Provider supply; healthcare demand
National Sample Survey of Nurse Practitioners (NSSNP)	NCHWA, HRSA, DHHS	https://bhwh.hrsa.gov/data-research/access-data-tools/national-sample-survey-registered-nurse-practitioners	Survey	Individual nurse practitioner	Provider supply

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis (cont.)

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
National Sample Survey of Registered Nurses (NSSRN)	NCHWA, HRSA, DHHS	https://www.census.gov/nssrn	Survey	Individual registered nurse	Provider supply
Uniform Data System (UDS)	HRSA, DHHS	https://bphc.hrsa.gov/datareporting/reporting/index.html	Establishment survey	Establishment	Establishment characteristics; provider staffing
Integrated Postsecondary Education Data System (IPEDS)	NCES, US Dept. of Ed.	https://nces.ed.gov/ipeds	Establishment survey	Establishment	Provider supply
National Mental Health Services Survey (N-MHSS)	SAMHSA, DHHS	https://www.samhsa.gov/data/data-we-collect/n-mhss-national-mental-health-services-survey	Establishment survey	Establishment	Establishment characteristics
National Survey of Substance Abuse Treatment Services (N-SSATS)	SAMHSA, DHHS	https://www.samhsa.gov/data/data-we-collect/n-ssats-national-survey-substance-abuse-treatment-services	Establishment survey	Establishment	Establishment characteristics; healthcare demand
National Survey on Drug Use and Health (NSDUH)	SAMHSA, DHHS	https://www.samhsa.gov/data/population-data-nsduh	Survey	Individual	Healthcare demand
American Community Survey (ACS)	Census	https://www.census.gov/programs-surveys/acs	Household survey	Household and individual	Provider supply
Current Population Survey (CPS)	Census, BLS, DOL	https://www.census.gov/programs-surveys/cps.html	Household survey	Individual	Provider supply
Economic Census	Census	https://www.census.gov/programs-surveys/economic-census.html	Establishment survey	Establishment	Establishment characteristics; provider staffing

TABLE A-1. Summary List of Federal Data Sources for Health Workforce Analysis (cont.)

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Relevance for Health Workforce Analysis
Survey of Business Owners and Self-Employed Persons (SBO)	Census	https://census.gov/programs-surveys/sbo.html	Company survey	Company	Company characteristics; provider staffing
Survey of Income and Program Participation (SIPP)	Census	https://www.census.gov/sipp	Household survey	Household and individual	Healthcare demand

LEGEND

AHRQ	Agency for Healthcare Research and Quality
BLS	Bureau of Labor Statistics
CDC	Centers for Disease Control and Prevention
Census	US Census Bureau
CMS	Centers for Medicare and Medicaid Services
DHHS	US Department of Health and Human Services
DOL	US Department of Labor
DRC	Data Resource Center for Child and Adolescent Health
US Dept. of Ed.	US Department of Education
NCHS	National Center for Health Statistics
NCES	National Center for Education Statistics
NCHWA	National Center for Health Workforce Analysis
HRSA	Health Resources and Services Administration
SAMHSA	Substance Abuse and Mental Health Services Administration



Health Resources & Services Administration

National Center for Health Workforce Analysis
Health Resources and Services Administration
5600 Fishers Lane
Rockville, MD 20857

Web: www.hrsa.gov



Center for Health Workforce Studies
School of Public Health, University at Albany
State University of New York
1 University Place, Suite 220
Rensselaer, NY 12144-3445

Phone: (518) 402-0250

Web: www.healthworkforceta.org

Email: info@healthworkforceta.org