

Compendium of Federal Data Sources to Support Health Workforce Analysis

April 2013



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



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

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

For more information about the National Center, e-mail us at healthworkforcecenter@hrsa.gov, or visit our website at <http://bhpr.hrsa.gov/healthworkforce/index.html>.

Table of Contents

Introduction	4
Table 1-Summary List: Federal Data Sources for Health Workforce Analysis	5
Agency for Healthcare Research and Quality (AHRQ)	
Healthcare Cost and Utilization Project (HCUP)	7
Medical Expenditure Panel Survey (MEPS)	9
Bureau of Labor Statistics	
Current Employment Statistics (CES) Survey	12
Employment Projections (EP) Program	13
Occupational Employment Statistics (OES) Survey	15
Census Bureau	
American Community Survey	17
Current Population Survey (CPS)	20
Centers for Disease Control and Prevention	
Behavioral Risk Factor Surveillance System (BRFSS)	22
National Ambulatory Medical Care Survey (NAMCS)	25
National Hospital Ambulatory Medical Care Survey (NHAMCS)	28
National Hospital Discharge Survey (NHDS)	30
National Study of Long-Term Care Providers (NSLTCP)	32
National Survey of Residential Care Facilities (NSRCF)	34
Centers for Medicare and Medicaid Services	
Medicare Claims Data	36
Medicaid Claims Data (Medicaid Analytic eXtract [MAX] data)	39
Medicare Current Beneficiary Survey (MCBS)	42
National Provider Identifier (NPI) File	45
Health Resources and Services Administration	
Area Resource File	47
National Center for Education Statistics	
Integrated Postsecondary Education Data System (IPEDS)	49

Introduction

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

This compendium provides a summary of 19 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 understand how these sources could be used for health workforce analysis. For each data source, this compendium provides the following:

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

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

For more information on the National Center, please go to the Center's website at: <http://bhpr.hrsa.gov/healthworkforce/index.html>, or email the Center at: Healthworkforcecenter@hrsa.gov.

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

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
Healthcare Cost and Utilization Project (HCUP)	Agency for Healthcare Research and Quality (AHRQ), U.S. DHHS	http://www.ahrq.gov/data/hcup/	health claims data	health care encounter	health care demand
Medical Expenditure Panel Survey (MEPS)	Agency for Healthcare Research and Quality (AHRQ), U.S. DHHS	http://meps.ahrq.gov/mepsweb/	household survey with component on sampled individuals health utilization data	individual	health care demand
Current Employment Statistics (CES) Survey	Bureau of Labor Statistics, US DOL	http://www.bls.gov/ces/ http://www.bls.gov/sae/	payroll/establishment survey (businesses and government agencies)	total employment by industry	total employment in health care establishments
Employment Projections (EP) Program (at Bureau of Labor Statistics)	Bureau of Labor Statistics, U.S. DOL	http://www.bls.gov/emp/	projection model	occupation	employment projections for health occupations
Occupational Employment Statistics (OES) Survey	Bureau of Labor Statistics, U.S. DOL	http://www.bls.gov/oes/	establishment mail survey	occupation	provider supply
American Community Survey (ACS)	U.S. Census Bureau, U.S. Department of Commerce	http://www.census.gov/acs/www/	household survey	household and individual	provider supply
Current Population Survey (CPS)	U.S. Census Bureau, U.S. Department of Commerce	http://www.census.gov/cps/	household survey	individual	provider supply
Behavioral Risk Factor Surveillance System (BRFSS)	Office of Surveillance, Epidemiology, and Laboratory Services, CDC, U.S. DHHS	http://www.cdc.gov/BRFSS/	telephone survey	individual	health care demand and access
National Ambulatory Medical Care Survey (NAMCS)	Division of Health Care Statistics, National Center for Health Statistics, CDC, U.S. DHHS	http://www.cdc.gov/nchs/ahcd.htm	office-based physician survey	health care encounter and physician	provider supply; health care demand
National Hospital Ambulatory Medical Care Survey (NHAMCS)	Division of Health Care Statistics, National Center for Health Statistics, CDC, U.S. DHHS	http://www.cdc.gov/nchs/ahcd.htm	establishment survey (hospital outpatient and emergency departments)	health care encounter	health care demand
National Hospital Discharge Survey (NHDS)	Division of Health Care Statistics, National Center for Health Statistics,	http://www.cdc.gov/nchs/nhds.htm	establishment survey (hospitals)	health care encounter	health care demand

Data Source	Federal Agency	Website	Data Collection Method	Unit of Analysis	Potential Use in Health Workforce Analysis
	CDC, U.S. DHHS				
The National Study of Long-Term Care Providers (NSLTCP)	Division of Health Care Statistics, National Center for Health Statistics CDC, U.S. DHHS	http://www.cdc.gov/nchs/nsltcp.htm	establishment survey	establishment	provider staffing; health care demand
National Survey of Residential Care Facilities (NSRCF)	National Center for Health Statistics (NCHS), Division of Health Care Statistics, CDC, U.S. DHHS	http://www.cdc.gov/nchs/nsrcf.htm	survey	individual health provider	Provider staffing; health care demand
Medicare Claims Data	Centers for Medicare and Medicaid Services, U.S. DHHS	http://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems.html	health claims data	health care encounter	provider supply; health care demand; health care access
Medicare Current Beneficiary Survey (MCBS)	Centers for Medicare and Medicaid Services, U.S. DHHS	http://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index.html	individual survey supplemented with claims data	individual	health care demand; health care access
National Provider Identifier (NPI File)	Centers for Medicare and Medicaid Services U.S. DHHS	http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProviderIdentifierStand/index.html	enrollment in program by health care providers	individual health provider	provider supply
Area Resource File (ARF)	National Center for Health Workforce Analysis, Health Resources and Services Administration, U.S. DHHS	http://arf.hrsa.gov/	compilation of multiple data sources to populate information on U.S. counties	county	provider supply
National Survey of Residential Care Facilities (NSRCF)	National Center for Health Statistics (NCHS), Division of Health Care Statistics, CDC, U.S. DHHS	http://www.cdc.gov/nchs/nsrcf.htm	survey	Individual health provider	Provider staffing; health care demand
Integrated Postsecondary Education Data System (IPEDS)	National Center for Education Statistics, U.S. Department of Education	http://nces.ed.gov/ipeds/	establishment survey	establishment	provider supply

Healthcare Cost and Utilization Project (HCUP)

FEDERAL AGENCY: Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services

WEBSITE: <http://www.ahrq.gov/data/hcup/>

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

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

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

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

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

patient characteristics and payment sources.

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

AVAILABILITY: Many of the HCUP databases are available for purchase through the HCUP Central Distributor (http://www.hcup-us.ahrq.gov/tech_assist/centdist.jsp). Information is also available using the tool, HCUPnet, which is a free, online query system based on data from the HCUP (<http://hcupnet.ahrq.gov/>). 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), U.S. Department of Health and Human Services

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

DESCRIPTION: The Medical Expenditure Panel Survey (MEPS), which began in 1996, is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States. The MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of health insurance held by and available to U.S. workers.

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

Household Component

The HC collects data from a sample of families and individuals in selected communities across the United States, drawn from a nationally representative subsample of households that participated in the prior year's [National Health Interview Survey](#) (conducted by the National Center for Health Statistics). During the household interviews, MEPS collects detailed information for each person in the household on the following: demographic characteristics, health conditions, health status, use of medical services, charges and source of payments, access to care, satisfaction with care, health insurance coverage, income, and employment. In the past several years, the sample size in the HC has ranged from about 12,000 to 14,000 families covering from about 30,000 to 35,000 individuals (http://meps.ahrq.gov/mepsweb/survey_comp/hc_sample_size.jsp). The 2010 Full

Year Consolidated File from the HC, the year available at the time of this writing, has 32,846 records.

Insurance Component

The IC collects data from a sample of private and public sector employers on the health insurance plans they offer their employees. The collected data include the number and types of private insurance plans offered (if any), premiums, contributions by employers and employees, eligibility requirements, benefits associated with these plans, and employer characteristics. 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 about 42,000 private sector establishments and 3,000 state and local governments.

Medical Provider Component (MPC)

MEPS also includes an MPC, which covers hospitals, physicians, home health care providers, and pharmacies identified by the MEPS HC respondents. Its purpose is to supplement and/or replace information received from the MEPS HC respondents. Data files containing only this supplemental respondent information are not available, but the information is incorporated into the MEPS HC data files.

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

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

levels.

AVAILABILITY: Summary tables with findings from MEPS are available at http://meps.ahrq.gov/mepsweb/data_stats/quick_tables.jsp. Downloadable public use files and documentation for the MEPS IC are available for data analysis at http://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, MD, restricted data files covering individuals and families, which otherwise are not publicly released for reasons of confidentiality (http://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp). For employer-level analyses, researchers with approved projects can access restricted data files covering business establishments through the U.S. Census Research Data Center network (<http://www.census.gov/ces/dataproducts/index.html>).

Current Employment Statistics (CES) Survey

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

WEBSITE: <http://www.bls.gov/ces/> and <http://www.bls.gov/sae/>

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

Each month the CES program surveys about 141,000 businesses and government agencies in the United States, representing more than 486,000 individual worksites.

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

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

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

Employment Projections (EP) Program

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

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

DESCRIPTION: The Employment Projections (EP) program at the BLS provides projections of the U.S. labor market for the following 10 years. The projections are updated every 2 years. BLS projections of industry and occupational employment are developed in a series of interrelated steps, each of which is based on a different procedure or model and assumptions. The methodology is described in the Chapter 13 of the BLS Handbook of Methods;

<http://www.bls.gov/opub/hom/pdf/homch13.pdf>; also see

http://www.bls.gov/emp/ep_projections_methods.htm.

The EP program uses survey data from several different sources to develop projections for the labor force, the macro economy, industry output and employment, and occupational employment. Data sources include the Current Population Survey, the Current Employment Statistics survey, the Occupational Employment Statistics survey, and the 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 (www.bls.gov/emp/#tables). Career information on the nature of the work, education and training requirements, job outlook, and more can be accessed from the Occupational Outlook Handbook (OOH) at www.bls.gov/ooh/healthcare/home.htm.

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

AVAILABILITY: Detailed projections for 300 industries and 750 occupations, as

well as summary tables on selected subjects, are available (<http://www.bls.gov/emp/#tables>). Detailed data files for researchers also are available (<http://www.bls.gov/emp/#data>). The OOH (www.bls.gov/oo/) also reports projections and other career information for more than 500 detailed occupations.

Occupational Employment Statistics (OES) Survey

FEDERAL AGENCY: Bureau of Labor Statistics, U.S. Department of Labor

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

DESCRIPTION: The Occupational Employment Statistics (OES) survey is primarily a mail survey measuring occupational employment and wage rates for wage and salary workers in nonfarm establishments nationally and, separately, in the 50 states and the District of Columbia, Guam, Puerto Rico, and the Virgin Islands.. An establishment is generally a single physical location at which economic activity occurs (e.g., store, factory, restaurant, hospital, physician office). The North American Industry Classification System (NAICS) is used to classify establishments by industry.

Respondents report their number of employees by occupation across 12 wage ranges. Self-employed workers, owners and partners in unincorporated firms, household workers, and unpaid family workers are not included in the survey. The Standard Occupational Classification (SOC) system is used to categorize occupations.

The sample size for reporting estimates of individual occupations varies by occupation, in part dependent on the prevalence of the occupation within the establishments sampled. Probability sample panels of about 200,000 establishments are selected semiannually. OES estimates are based on six panels of establishment survey data and rely upon data from approximately 1,200,000 sampled establishments collected over a 3-year period. (See http://www.bls.gov/oes/current/methods_statement.pdf for more information.)

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

Employment and wage information for an occupation can be detailed by industry.

Thus, for example, information is available on the employment and wages of registered nurses nationally in general hospitals, physician offices, home health, nursing facilities, and outpatient centers, among others (e.g., <http://www.bls.gov/oes/current/oes291111.htm#ind>).

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

AVAILABILITY: Summary tables are available on the OES website. Downloadable Excel spreadsheets and text files with occupational employment and wage estimates are also available (see http://www.bls.gov/oes/oes_dl.htm and <ftp://ftp.bls.gov/pub/time.series/oe/>).

American Community Survey

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

WEBSITE: <http://www.census.gov/acs/www/>

DESCRIPTION: The American Community Survey (ACS) is an ongoing statistical survey that samples a small percentage of the population in the United States and Puerto Rico every year. The ACS contains a sample of housing units and group quarters.

The ACS includes information on age, sex, race and ethnicity, language, disability, health insurance status, state of residency and employment, employment status, hours worked, occupation, education, income, household size and characteristics, and family characteristics and relationships, among other items.

There are three types of ACS data files: 1-year files, 3-year files, and 5-year files. The 1-year file contains data for a single year, the 3-year file combines data from 3 years, and the 5-year file combines 5 years of ACS data.

The ACS annually collects data from about 2 million households. The sample is based upon the Census Bureau's official inventory of known living quarters, the Master Address File. Data are collected on all individuals in a sampled household and on a sample of individuals in sampled group quarters. About 1 in 38 households per year receives an invitation to participate in the ACS.

The 2011 ACS public use microdata sample files (PUMS) represent about 1 percent of households: there are 1,338,380 housing unit records and 2,982,598 person records from households and 163,204 person records from group quarters (http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2011AccuracyPUMS.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, hours worked—in addition to demographic information on age, sex, race and ethnicity, et cetera—can be used to

estimate the size and characteristics of a specific health occupation. Occupation classification coding schemes include the Standard Occupational Classification (SOC) scheme and the Census classification scheme (based upon SOC) (see, e.g., http://www.census.gov/acs/www/Downloads/data_documentation/CodeLists/2011_ACS_Code_Lists.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 on individual demographics, health insurance status, and disability status may be applicable to studies of health care needs underlying demand for health care providers.

GEOGRAPHICAL DETAIL: The geographical detail in estimates published in summary form by the Census Bureau varies based on whether estimates are from a 1-year, 3-year, or 5-year file. The reliability of estimates for geographical units with the smallest population numbers increases as the number of years combined in a file increases. The Census Bureau reports summary estimates for areas with populations of 60,000 or more from 1-year files, 20,000 or more from 3-year files, and on geographic areas with fewer than 20,000 with 5-year files.

In the PUMS, which contain individual household- and person-level data, the smallest geographical units of analysis are public use microdata areas, which are areas with a population of at least 100,000 persons.

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

The ACS summary data are presented in predefined tabulations of characteristics. The basic unit of analysis is a specific geographic entity—for example, a state or county—for which estimates of persons, families, households, or housing units in particular categories are provided. A user can select specific predefined tabulations through the Census Bureau's American Fact Finder (<http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>). Summary data

files provide access to the detailed tables through a series of comma-delimited text files on the Census Bureau's File Transfer Protocol site of the tabulations.

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

Information on downloading PUMS and summary data files from the ACS is available at http://www.census.gov/acs/www/data_documentation/data_via_ftp/.

Handbooks for ACS data users are available at http://www.census.gov/acs/www/guidance_for_data_users/handbooks/.

Details on the ACS survey methodology and content are available at http://www.census.gov/acs/www/methodology/methodology_main/.

Current Population Survey (CPS)

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

WEBSITE: <http://www.census.gov/cps/>

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

The CPS is administered by the Census Bureau using a monthly probability sample of occupied housing units. The fieldwork is conducted during the calendar week that includes the 19th of the month. The questions refer to activities during the prior week; that is, the week that includes the 12th of the month. Households from all 50 states and the District of Columbia are in the survey for 4 consecutive months, are out of the survey for 8 months, and then return for another 4 months before leaving the sample permanently. A personal visit interview is required for all first month-in-sample households. For households in the sample for the second, third, and fourth months, the preferred method for data collection is a telephone interview. The preferred data collection method for the household's fifth-month interview is a personal interview. This fifth-month interview follows a sample unit's eight-month dormant period. Fifth-month households are more likely than any other sampled households to be ones in which the previous residents have moved out and been replaced by an entirely different group of residents.

In addition to the regular labor force questions, the CPS often includes supplemental questions on subjects of interest to labor market analysts. These include annual work activity and income, veteran status, school enrollment, contingent employment, worker displacement, and job tenure, among other topics.

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

150,000 person-level records in each monthly file (see http://thedataweb.rm.census.gov/ftp/cps_ftp.html).

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The CPS has data elements relevant to analyses of health workforce supply and labor force participation. Variables such as occupation, employment status, 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 on the CPS microdata file for the main job and any second job a person held during the reference week of the data collection. Currently the codes represent the Census 2010 classification system, which was developed using the Standard Occupational Classification (SOC) Manual: 2010. The CPS may also enable analysis of trends in health workforce supply because it is conducted monthly. The empirical trends derived from the CPS may, for example, illuminate the rate of new entrants in 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 variables that indicate the metropolitan status of a participant's residence, e.g., metropolitan or nonmetropolitan.

AVAILABILITY: Basic monthly CPS data files and the annual March or Annual Social and Economic Supplement data files, along with documentation, can be downloaded from http://thedataweb.rm.census.gov/ftp/cps_ftp.html. Summary data tables from the annual supplement data are available at <http://www.census.gov/cps/data/>. Summary data files and tables on labor force statistics are also available from the BLS at <http://www.bls.gov/cps/home.htm>.

More detail on the CPS methodology is available at <http://www.census.gov/cps/methodology/>.

Behavioral Risk Factor Surveillance System (BRFSS)

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

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

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

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

Before 2011, the sampling frame for the survey was based on lists of traditional landline telephone numbers. In 2011, 50 states, the District of Columbia, Guam, and Puerto Rico collected samples of both landline and cell phone interviews while the Virgin Islands collected a sample of landline-only interviews. Information on the methodological changes associated with the 2011 BRFSS and the potential effects of these changes on prevalence estimates is available at <http://www.cdc.gov/surveillancepractice/reports/brfss/brfss.html>.

The 2011 BRFSS file has about 500,000 records, the 2010 BRFSS more than 450,000 records, and the 2009 BRFSS more than 430,000 records.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: BRFSS is one potential source of data on health care utilization and access of the adult population, used to gauge demand for services. It is not intended—in sample design or purpose—for

analysis of the supply of providers. The BRFSS data collection enables estimates of health status (including prevalence of chronic conditions), health utilization, and health access.

GEOGRAPHICAL DETAIL: State-level estimates and some county- and metropolitan/micropolitan-level estimates are available from the annual BRFSS public use file. Although there is a variable for identifying the county of a respondent, this field is often missing and not reported in the public use file to protect the confidentiality of respondents. Hence, identification of 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 pre-tabulated form through the BRFSS SMART (Selected Metropolitan/Micropolitan Area Risk Trends) (see <http://apps.nccd.cdc.gov/BRFSS-SMART/>).

AVAILABILITY: Annual public use BRFSS data files incorporating the monthly data collected by each state are available from the CDC. Data files and documentation are available at http://www.cdc.gov/brfss/technical_infodata/surveydata.htm. These data files enable primary data analysis on the case-level data by analysts.

The BRFSS WEAT (Web Enabled Analysis Tool) enables users to create cross tabulation analysis and logistic regression analysis from variables available in the WEAT variable lists (see http://apps.nccd.cdc.gov/s_broker/weatsql.exe/weat/index.hsql). The analysis selected by the user is run through the WEAT system with the generated results, reported on the WEAT website, downloadable as an Excel spreadsheet. A suppression protocol is applied, i.e., estimates are suppressed when the denominator sample size is less than 50.

As noted, pre-tabulated data tables and charts for selected

metropolitan/micropolitan areas are available at <http://apps.nccd.cdc.gov/BRFSS-SMART/>. SMART data files and documentation are available at http://www.cdc.gov/brfss/smart/technical_infodata.htm. These data files enable primary analysis on the case-level data by analysts.

National Ambulatory Medical Care Survey (NAMCS)

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

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

DESCRIPTION: The National Ambulatory Medical Care Survey (NAMCS) is an annual national survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States. The survey was conducted annually from 1973 to 1981, in 1985, and annually since 1989.

The NAMCS is based on a sample of practicing physicians and a sample of patient visits to each sampled physician's office-based practice. Each sampled physician is randomly assigned to a 1-week reporting period. During this period, data for a systematic random sample of visits are recorded. Data are obtained on patients' symptoms, physician diagnoses, and medications ordered or provided. The survey also provides statistics on the demographic characteristics of patients and services provided. Data are also intermittently collected on special topics, such as a series of questions in 2012 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 (2006-10), electronic health record/electronic medical record (EHR/EMR) (2008-13), and physician workflow (2011-13). The EHR/EMR and Workflow supplements pertain to the adoption, benefits and barriers of electronic health records in physician office-based practices. Both supplements were sponsored by the Office of the National Coordinator for Health Information Technology, U.S. Department of Health and Human Services. A "lookback" module was added in 2012 that collects, on the sampled visit, clinical data on the previous twelve months. This module is triggered when certain cardiovascular diagnoses are made in order to evaluate and monitor the appropriateness of clinical management and

the relationship to these outcomes.

Through the 2011 survey year, an annual NAMCS data file has typically included about 30,000 (visit) records based on data collected from between 1,200 to 1,400 sampled physicians. Most recently in 2012 and 2013, the number of office-based physicians in the drawn sample has been increased to between 15,000 and 20,000 which will result in a larger visit-level data file. Since 2006, the NAMCS also has sampled providers in community health centers (CHC), 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.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: NAMCS data have been used to make estimates about the volume and type of ambulatory care visits at physician offices. This type of information may be used to estimate demand for ambulatory care providers by describing the visit rates and patterns across 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 public use files from 2005 forward contain a physician-level weight that can be used to make estimates of office-based physicians.

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

GEOGRAPHICAL DETAIL: National and regional estimates can be made with sufficient reliability from historical public use data files. The 2012 increase in sample size is intended, in part, to enable estimates for a majority of the states in the United States.

AVAILABILITY: Documentation on NAMCS data files and downloadable public use files are available from http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm.

Restricted data files, which include some data elements not available in public use files, are accessible for approved projects through the Research Data Center at the National Center for Health Statistics (<http://www.cdc.gov/rdc/>).

National Hospital Ambulatory Medical Care Survey (NHAMCS)

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

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

DESCRIPTION: The National Hospital Ambulatory Medical Care Survey (NHAMCS) is designed to collect data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments. Findings are based on 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, exclusive of Federal, military, and Veterans Administration hospitals in the 50 states and the District of Columbia. Within emergency service areas, outpatient department clinics, or ambulatory surgery locations, patient visits are systematically selected over a randomly assigned 4-week reporting period.

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

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: NHAMCS is designed to provide estimates of visits to EDs, OPDs, and ASLs. It is not designed for estimates of provider supply within EDs, OPDs, and ASLs. The information in NHAMCS may be used to estimate 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 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 are available from the website at http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm. Restricted data files, which include some data elements not available in public use files, are accessible for approved projects through the Research Data Center at the National Center for Health Statistics (<http://www.cdc.gov/rdc/>).

National Hospital Discharge Survey (NHDS)

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

WEBSITE: <http://www.cdc.gov/nchs/nhds.htm>

DESCRIPTION: The National Hospital Discharge Survey (NHDS), which was conducted annually from 1965-2010, was a national probability survey designed to meet the need for information on characteristics of inpatients discharged from non-Federal, noninstitutional, short-stay hospitals in the United States. Two data collection procedures were used. One was a manual system in which sample selection and transcription of information from hospital records to abstract forms were performed by the hospital's staff or by staff of the U.S. Census Bureau on behalf of NCHS. The other was an automated system in which NCHS purchased computer files containing electronic data files from commercial organizations, state data systems, hospitals, or hospital associations. Records from these files were systematically sampled by NCHS. Approximately 45 percent of the respondent hospitals provided data through the automated system.

The NHDS will be 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 beginning 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) varies 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.

AVAILABILITY: Downloadable public use data files and documentation are available with no charge at http://www.cdc.gov/nchs/nhds/nhds_questionnaires.htm. Restricted data files, which include some data elements not available in public use files, are accessible for approved projects through the Research Data Center at the National Center for Health Statistics (<http://www.cdc.gov/rdc/>).

National Study of Long-Term Care Providers (NSLTCP)

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

WEBSITE: <http://www.cdc.gov/nchs/nsltcp.htm>

DESCRIPTION: The National Study of Long-Term Care Providers (NSLTCP) is a new study, implemented in 2012. The purpose of the NSLTCP is to integrate all existing and future long-term care (LTC) provider surveys at NCHS into a unified study. As such, the NSLTCP replaces the previous National Nursing Home Survey and the National Home and Hospice Care Surveys conducted at NCHS.

The NSLTCP will be conducted every two years to provide data on an array of LTC settings. The inaugural survey will include data on residential care facilities and adult care service centers in addition to nursing homes, home health agencies, and hospices that were the subjects of the earlier two studies. The NSLTCP will collect primary data to obtain information on residential care facilities and adult care service centers where there is a paucity of nationally representative administrative data on these settings. Existing administrative data on nursing homes, home health agencies, and hospices will be brought together into a unified dataset of comparable information. The NSLTCP enables comparisons across different LTC provider types within a similar time period.

The sample for the inaugural survey components of NSLTCP includes about 5,000 adult care service centers and 11,700 assisted living and other residential care communities. For the administrative data, the universe of nursing homes, home health care agencies, and hospices available through the data sources will be included.

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NSLTCP includes data on the numbers and types of staff across LTC providers, in particular the number of 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: The long-range objective is to have coverage in the data collection to enable national and state estimates, where feasible, on the supply, use, and characteristics of LTC establishments encompassed in the study, which represent the major provider sectors of the U.S. LTC system.

AVAILABILITY: The first public use survey data files from the NSLTCP are expected to be released in 2014, the first reports in late 2013, and the first survey data files (available through the NCHS Research Data Center) in the summer of 2013.

National Survey of Residential Care Facilities (NSRCF)

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

WEBSITE: <http://www.cdc.gov/nchs/nsrcf.htm>

DESCRIPTION: The National Survey of Residential Care Facilities (NSRCF) is the first-ever national survey of assisted living and other residential care providers and their residents. The NSRCF was designed to produce national estimates of these places and residents.

Included are: residential care facilities; assisted living residences; board and care homes; congregate care; enriched housing programs; homes for the aged; personal care homes; and shared housing establishments that are licensed, registered, listed, certified, or otherwise regulated by a state. Facilities eligible for this study are residential care facilities, defined as places that were: licensed, registered, listed, certified, or otherwise regulated by a state; had four or more licensed, certified, or registered beds; provided room and board with at least two meals a day and around-the-clock on-site supervision; helped with personal care such as bathing and dressing or with health-related services such as medication management; and served a predominantly adult population. Facilities licensed to serve the severely mentally ill or the developmentally disabled populations exclusively, and facilities that did not have any current residents, were excluded. Nursing homes were also excluded unless they had a unit or wing meeting the definition outlined above and had residents could be separately enumerated.

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

RELEVANCE FOR HEALTH WORKFORCE ANALYSIS: The NSRCF provides data in two separate files. The provider/facility file includes data on provider characteristics, such as physical structure and environment; types of services offered; types of staff employed; benefits for, training of, and job roles by personal

care aides; and policies on admission, retention, and discharge. The resident file includes data on resident characteristics, such as 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 NSRCF is to provide national estimates of: (1) the number of residential care facilities operating in the United States; (2) the number of residents receiving care; and (3) the characteristics of both the facilities and their residents.

AVAILABILITY: The public use data files were released in Dec 2011, along with a data brief describing first findings and a methods report. Data files can be downloaded from

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010

Before downloading the files, data users should review the information at

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_FacilityFileDataReadme.pdf and

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NSRCF/2010/2010NSRCF_ResidentsFileDataReadme.pdf.

Medicare Claims Data

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

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

DESCRIPTION: The Medicare claims files contain information collected by Medicare to pay for health care services provided to a Medicare beneficiary. Data are available for each institutional and non-institutional claim type, with each record being a claim. Some of the information contained within these utilization files includes: procedure and diagnosis information, dates of service, revenue center detail, payment and charge amounts, beneficiary demographic information, and limited professional provider and facility data. Each Medicare claims file contains information for a calendar year.

Claim or file types are: skilled nursing facility; outpatient; inpatient; hospice; home health agency; durable medical equipment; and carrier (also known as the physician/supplier Part B claims file). See <http://www.resdac.org/cms-data/file-family/Medicare-Claims>.

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

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

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

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

AVAILABILITY: Medicare claims data are available in research identifiable files (RIF), limited data set (LDS) files and, recently, public use files (PUF).

RIFs include beneficiary-level protected health information and, thus, requests for RIF data require a Data Use Agreement (DUA) and review by CMS' 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 description by the Research Data Assistance Center at <http://www.resdac.org/resconnect/articles/148> for more information about RIFs,

LDS files, and PUFs.

RIFs are available as custom files or 5 percent sample files. For more information on Medicare RIFs and how to request these files see

<http://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/IdentifiableDataFiles/index.html>.

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

Basic Stand Alone Medicare claims PUFs can be accessed and downloaded directly from the CMS website at http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/BSAPUFS/index.html?redirect=/BSAPUFS/10_SNF_Bene_PUF.asp.

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

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

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

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

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

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

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

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

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

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

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

Medicare Current Beneficiary Survey (MCBS)

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

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

DESCRIPTION: The Medicare Current Beneficiary Survey (MCBS) is a multipurpose survey of a nationally representative sample of Medicare beneficiaries. The MCBS uses a rotating panel design, in which four panels of about 4,000 participants each are active at a given time. A new panel is introduced each year in the fall and replaces the oldest panel. The MCBS data contain information on socioeconomic and demographic characteristics, health status and functioning, health care use and expenditures, health insurance coverage (including Medicare, Medicaid, and private insurance), and Medicare claims.

The MCBS has two data modules: 1) the Access to Care module and 2) the Cost and Use module. The Access to Care file contains information on beneficiaries' access to health care, satisfaction with care, and usual source of care. (Information collected in the MCBS is augmented with data on the use and program cost of Medicare services from Medicare claims data.) The Cost and Use file links Medicare claims to survey-reported events and provides complete expenditure and source-of-payment data on all health care services, including those not covered by Medicare. MCBS-reported data include information on supplementary health insurance, living arrangements, income, health status, and physical functioning, among others.

A detailed description of the differences between the Access to Care module and the Cost and Use module is available at the Research Data Assistance Center (ResDAC) at <http://www.resdac.org/resconnect/articles/103>. A section of the ResDAC description is quoted below:

The Access to Care module will contain four panels of participants – the new panel plus the three continuing panels. The Cost and Use

module contains only the three continuing panels. The Access to Care module is designed to represent the “always enrolled” Medicare population. This means that this module will not include individuals who become enrolled after the first day of the year or individuals who disenrolled or died prior to the last day of the year. The Cost and Use module, conversely, is designed to represent the “ever enrolled” Medicare population and does not apply those exclusions. This distinction is important, particularly for those who are interested in studying mortality—Cost and Use is the appropriate module to request for that purpose.

The Access to Care module has a sample of about 16,000 (four panels of 4,000) and the Cost and Use module has a sample of around 12,000 (three panels of 4,000).

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

GEOGRAPHICAL DETAIL: The MCBS includes a variable indicating the U.S. Census Bureau geographic division or region.

AVAILABILITY: All research requests for MCBS data must be processed through ResDAC at www.resdac.umn.edu or via email at resdac@umn.edu. Once a data request has been reviewed by ResDAC, a formal request package is sent to CMS.

The MCBS is currently \$600 per module per year.

CMS prepares and releases sourcebooks or data tables reporting findings from the MCBS, available at <http://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables.html>.

National Provider Identifier (NPI) File

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

WEBSITE: <http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/index.html>

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

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

numbers, or dates of birth.

Starting with the October 2012 file (released in November 2012), the file includes cases in which the NPI number has been deactivated and the deactivation date. The reason for deactivation is not included.

The NPI file contains data on all providers who obtained a NPI number. The file as of October 2012 had over 3.8 million records with active NPI numbers—including 2.9 million for individual providers and 900,000 for organizations.

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 address in the NPI file, which includes ZIP code, offers the opportunity for geographical coding to small areas.

AVAILABILITY: The full public use data file, which is updated monthly, can be downloaded at http://nppes.viva-it.com/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

<http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/DataDissemination.html>.

Area Resource File

FEDERAL AGENCY: Health Resources and Services Administration, Bureau of Health Professions, National Center for Health Workforce Analysis, U.S. Department of Health and Human Services

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

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

The ARF entails no sampling, as it contains data on all the counties in the United States, which number more than 3,000.

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

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

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

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

AVAILABILITY: ARF data can be downloaded at no cost, or a CD can be purchased at a nominal price. The data files are available at <http://arf.hrsa.gov> or <http://datawarehouse.hrsa.gov/arf.aspx>.

Integrated Postsecondary Education Data System (IPEDS)

FEDERAL AGENCY: National Center for Education Statistics, U.S. Department of Education

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

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

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

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

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

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

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

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