As implementation of health reform initiatives accelerate, states are increasingly aware of the need to collect state-level health workforce data to determine adequacy of the current health workforce to meet the expected increase in demand for health services. This brief is intended to assist states and organizations who are engaged in health workforce planning by answering some frequently asked questions about health workforce data collection.*

Q: Where can I find health workforce data and related data?

A: There are many sources of health workforce data. Some sources have known and documented limitations. It is important to understand the data's purpose and how they are collected, verified, and updated.

There are 2 reports that describe multiple data sources:

- The [Health Workforce Analysis Guide, 2016 Edition](#) lists selected federal, nonfederal, and state data sources, and has a brief section on general data limitations.
- The [Compendium of Federal Data Sources to Support Health Workforce Analysis](#) contains information on individual data sets such as NPPES, MEPS, and NHANES. Some of the sources included in the Compendium are freely available. Others require an application, data use agreement, and/or payment.

*Selected federal sources:*

The [Bureau of Labor Statistics](#) (BLS) is a commonly used federal data source included in the compendium. The BLS tracks employment by industry and occupations, projects future employment, houses the Current Population Survey, and provides other employment statistics. A known limitation is that the [BLS Occupational Employment Statistics](#) data count jobs, not workers, and excludes workers who are self-employed, unemployed, or in certain industries. Professions with a large number of self-employed workers, such as physicians and dentists, may be underestimated, while professions with workers that work 2 or more part-time jobs, such as [dental hygienists](#), may be overestimated. The Current Population Survey, which surveys households, is another commonly-used dataset from BLS that is used to estimate health workforce statistics.

The [Area Health Resource File](#) (AHRF) is a publicly available dataset that aggregates data from disparate data sources. It contains county-level and state-level data on healthcare workers and other demographic and health-related variables. Some variables based on data from the American Dental Association, the American Hospital Association and the American Medical Association are subject to copyright restrictions.

*Selected nonfederal sources:*

Professional associations, such as the [American Medical Association](#) (AMA), the [American Dental Association](#) (ADA), and the [American Hospital Association](#) (AHA) conduct their own surveys and maintain databases (eg, “Masterfiles”) for administrative and analytic uses. These data sources are often proprietary and available for purchase under strict data use agreements.

* Web resources are hyperlinked throughout the document and listed in full at the end.
SK&A maintains databases on physicians and other healthcare practitioners. They claim that the lists are verified every 6 months and updated monthly, and that mailing lists are guaranteed 100% deliverable.

Another nonfederal data source is Kaiser Family Foundation State Health Facts. Like the AHRF, this source reports aggregated data on providers, service use, and other useful health-related variables obtained from outside resources.

State sources:

States may or may not collect their own workforce-related data. The State Health Workforce Data Collection Inventory lists states that collect data on health workforce supply, demand, and education.

Q: Why don’t different data sources match?

A: There are multiple approaches to collecting data, and data are often collected for different purposes. As a result, it is important to understand the methodology behind each dataset and its intended use in order to make valid comparisons. For example, the Bureau of Labor Statistics Occupational Employment Statistics collects data from employment surveys; the data count jobs, not workers, and they count employed, not self-employed positions. Professional association masterfiles (e.g., AMA, ADA) are based on membership surveys and other sources, and data may not accurately account for professionals that are licensed and practicing in more than one state. State licensure data are self-reported through license applications and renewals, and hinge on the licensees' accuracy and timeliness. The National Provider and Plan Enumeration System (NPPES) is a registry of providers that submit Medicare and Medicaid claims; this is an administrative database where the billing address of the provider may not match the provider’s practice location. Health professionals are mobile, some more than others, and change jobs and locations; these moves may not be reflected accurately or in a timely manner.

Q: How reliable are different sources of physician supply data?

A: An article by DesRoches et al (2015) compared the National Provider and Plan Enumeration System (NPPES), the American Medical Association Masterfile, and the SK&A physician file to evaluate data accuracy. The authors performed this analysis in the context of using the selected datasets for sampling frameworks and counting physicians in a given area. The authors found that while none of the files were perfect, the NPPES contained broader coverage and NPPES and SK&A data had reasonably accurate and current address information. The AMA Masterfile had lower rates of correct address information.

State licensure data are another matter. Some state medical boards require only basic information, including a mailing address for licensing correspondence. Some states collect more robust data through licensure, including multiple practice addresses, and demographic, education, and practice characteristics. Some states conduct regular surveys. States may or may not systematically verify the licensure or survey data.

Different data sources have different limitations. Before using any dataset as a sampling frame or for research, it is essential to understand the data's purpose and how they are collected, verified, and updated.

Q: What is the MDS?

A: The Minimum Data Set, or MDS, provides basic, consistent guidelines for fundamental health workforce questionnaires. These questions can be used by anyone who wants to collect data on the supply of health workers, whether through the licensure process or surveys, and can be adapted for additional professions. MDS questions focus on essential demographic, education, and practice characteristics. For more information, click here.
Q: What states have implemented the MDS?

A: Many states are already collecting health workforce data, with a customized MDS in place to collect any additional data they need for health workforce planning. Some examples of states that are already collecting an MDS include North Carolina, Virginia, New York, Indiana, and Minnesota.

For more information on which states are collecting data, visit our inventory page or contact HWTAC.

Q: Do I have to ask questions as recommended by the Minimum Data Set (MDS) or can I ask for other information?

A: The Minimum Data Set (MDS) provides guidelines for collecting basic, minimum, and consistent data on health professionals. These guidelines are not requirements, but they do provide suggestions so that data are collected in a way that is useful for research purposes and comparable across professions and states. Some states ask questions that go beyond the MDS so they can better understand their workforce and answer questions from their policymakers.

The following resources provide information on basic MDS guidelines and going beyond the MDS to ask additional questions, plus examples of data collection instruments from various states.

- Minimum Data Sets Created by National Boards, Councils, and Associations
- The Health Workforce Minimum Data Set (MDS): What You Need to Know
- Beyond the Minimum Data Set (MDS): What Additional Data Do States Collect on Health Workforce Supply?
- State Health Workforce Data Collection Inventory (look for states with “Yes” under “Surveys Available”)

Q: What are some different ways to collect health workforce data?

A: There are generally 4 methods to collect health workforce data:

1. **Licensure Process.** Data are collected as part of the licensure process when health professionals apply for their initial license and when they renew, capturing 100% of the workforce. This is one of the most efficient and cost-effective methods to collect data. Some questions on the licensure forms may be mandatory, while others are optional. The organizational structure of the licensing boards will present different opportunities and barriers to data collection. Examples: North Carolina, South Carolina, Virginia

2. **Surveys.** Data are collected through surveys, either in conjunction with the licensure process or as a separate effort. This method requires more staff time and money. Response rates may vary, but this is a good option if health workforce questions cannot be included directly on the licensure forms. Examples: New York, Wisconsin

3. **Continuous Monitoring.** Data collection begins with a list of all licensees in one or more professions. From there, states track individuals through surveys, news clipping services, and other methods to determine practice status, practice setting, and other characteristics. This method can be costly, but it may provide more up-to-date information. Examples: Iowa, Nebraska

4. **Secondary Data Sources.** Secondary data sources can also be used to enumerate the workforce in a specific state. These data sources include the National Provider Identification (NPI) file, the American Medical Association (AMA) Physician Masterfile, the US Bureau of Labor Statistics, and the Census Bureau’s American Community Survey, as well as state professional associations. Additionally, all-payer claims databases can be used to enumerate the health workforce in select states, but there are significant limitations.
Q: How do you work with licensure boards to collect and share data?

A: Relationships are key. Licensure boards are important partners in health workforce data collection, but their main priority is regulation to protect patient safety. They often don't have resources (i.e., funding, staff, time) to collect additional data, and in some states, current legislation restricts their ability to share data.

Show the boards the value of collecting additional workforce data as it relates to evidence-based regulation, and look for ways to minimize their burden, especially during the initial development period. Treat them as a valued partner and bring them into the conversation very early to build trust.

- Collaborating with Licensing Bodies in Support of Health Workforce Data Collection: Issues and Strategies

Q: How easy is it to get licensure boards to add or change questions?

A: This will vary from state to state. It is important to remain cognizant of a) the financial cost to the board to change online renewal questions; b) the time that it takes respondents to complete their licensure renewal form; and c) the need for comparability across time. Only request changes or additions when absolutely necessary.

Some states mandate the collection of data through legislation, which affects how easy it is to add or change questions. For example, Florida's data collection is legislated, and any question must go through a lengthy public comment period to be added or changed. This process has the potential to subject questions to bias from the public and special interest groups.

Q: Do you have examples of questions that we could ask?

A: Yes. The National Forum of State Nursing Workforce Centers, and the Federation of State Boards of Physical Therapy (FSBPT) have developed Minimum Data Set questions for their professions. Additionally, HRSA has developed MDS standards, and the WWAMI Center for Health Workforce Studies at the University of Washington has archived a questionnaire library containing data collection instruments volunteered by several states. The HWTAC is also including selected data collection instruments in the State Health Workforce Data Collection Inventory.

- National Center for Health Workforce Analysis, HRSA
- National Forum of State Nursing Workforce Centers
- FSBPT
- WWAMI Center for Health Workforce Studies

Q: What are the best rural definitions to use for health workforce analysis? Where can I find them?

A: Multiple rural definitions can be used in health workforce analysis. The study's purpose, unit of analysis, and funder should drive which rural definition is used.

The WWAMI Rural Health Research Center at the University of Washington is a leading resource on analyzing the rural health workforce. See Chapter 3 in their 2003 report, State of the Health Workforce in Rural America: Profiles and Comparisons for guidance on strengths and weaknesses of common rural definitions.

The Rural Health Information Hub, or RHIhub, is another important resource on rural definitions. The RHIhub developed an “Am I Rural?” tool that helps determine if a specific location is considered rural, including definitions used in federal program eligibility criteria. Additionally, states may have their own definitions of rural.
Q: What's the best geographic unit to use for health workforce analysis?

A: The study's purpose, design, data, confidentiality considerations, and funder requirements should inform which geographic unit of analysis is most appropriate. For a broader discussion of geographic units, see Chapter 5, Geography and Disparities in Health Care (Ricketts 2002) in the IOM's Guidance for the National Healthcare Disparities Report.

Studies that work with small cell sizes, especially in small geographic units, should consider the risk of deductive disclosure, where an individual's identity may be ascertained using known characteristics, such as race and age, even when direct identifiers, such as name and address, are removed.

Studies that use multiple datasets at varying geographic units can aggregate or approximate the data to larger geographic areas. For example, the FutureDocs Forecasting Tool created tertiary service areas by aggregating groups of counties to approximate the Dartmouth Atlas Hospital Referral Regions, which are built on ZIP Codes.

Analyses using sample data should evaluate the data's sampling unit and frame when determining an appropriate geographic unit of analysis. Data derived from a sample will have some degree of uncertainty associated with the estimates. Generally, the smaller the sample the larger the sampling error. Relying on a small geographic unit may further exacerbate the uncertainty around the estimates and prevent researchers from producing reliable statistics. Therefore one should carefully consider the data-generating process before considering the geographic unit of analysis.

Q: What are some new directions that health workforce research and planning are taking?

A: While it is important to understand how many health professionals there are and in which professions, specialties, employment settings, and geographic locations they practice, health workforce research is moving beyond understanding supply to better understanding demand for health professionals, how they are training and practicing, how they impact the quadruple aim, and how to more effectively plan for the future. The Global Health Workforce Alliance reports, “The current discourse on HRH is evolving from an exclusive focus on availability of health workers – i.e. numbers – towards according equal importance to accessibility, acceptability, quality and performance.”

A special issue of Health Services Research in 2017 provides a summary and examples for how health workforce research is evolving. Washko and Fennell summarize 4 main themes, including “(1) the changing roles of health care providers, (2) the changing combinations of different providers who work together to deliver care, (3) the impact of these workforce changes on quality of care and access to care, and (4) advances in methodological challenges inherent in the study of evolving health workforce changes.”

HRSA has funded 7 Health Workforce Research Centers (HWRCs) to conduct and disseminate “rigorous research that strengthens evidence-based policy and enhances government’s and the public’s understanding of issues and trends in the health workforce” to help inform health workforce planning and policy. The HWRCs’ research focuses on allied health, behavioral health, “emerging topics”, long-term care, oral health, and technical assistance. In 2017, the George Washington University HWRC compiled a report, Health Workforce Centers (HWRCs) Key Findings, 2013-2016, that identifies 3 main themes in the HWRCs’ work, including understanding the evolving health workforce configuration; spotlighting job growth and career paths in middle- and low-skilled health professions; and identifying workforce strategies to increase access to high-quality health care.
Q: How do you measure demand for health workers?

A: Demand for health services can be difficult to measure, and data availability varies. Broadly speaking, demand for health services can be split into 2 categories:

- **Utilization**
  - Those who utilize health care services, which includes people who need and receive services, and people who receive but may not need services (eg, elective procedures, the “worried well”)

- **Unmet Need**
  - Those who need services but do not choose to seek them
  - Those who need services but cannot access them because of limiting factors such as cost, insurance coverage, time, transportation, availability of healthcare providers, or other reasons

Utilization can be measured by claims data and sample surveys such as the Medical Expenditures Panel Survey (MEPS), but this underestimates demand for services. Data measuring unmet need is not systematically collected, and thus must be estimated or captured through individual surveys.

From the supply side, job vacancy, turnover, recruiting bonuses, and employment projections are also indicators of demand for health care services and workers. The Bureau of Labor Statistics tracks changes in employment and projects future employment estimates. Job vacancy data can be tracked through job boards or proprietary data sources such as Burning Glass Industries. Other vacancy, turnover, and bonus data can be tracked through hospital and other industry surveys. An example of state-level demand tracking is the Washington Health Workforce Sentinel Network. The Sentinel Network links health care employers with educators, policymakers and workforce planners to identify and respond to new and changing demand for healthcare workers, skills and roles.

Patient population factors, such as aging of the population, and policy changes that affect insurance coverage and disease burden, also influence future estimates of demand.

Q: What are the best ways to communicate and disseminate research and data to inform policy?

A: Stakeholders engaged in legislative, education, practice, payment, and regulatory policy discussions need data to help inform their decisions. Data should be presented in different formats (eg, briefs, slides, fact sheets) and at different levels (eg, academic research vs layperson language) depending on the audience. Some health workforce researchers have been advised to cultivate connections with legislative aides and to communicate research findings through social media. Others have been advised to present their data and research on a single page in short, concise bullets and easy-to-read graphics. It is important to highlight key messages and minimize less useful information.

For additional resources on using data and research to inform policy, see:

- State Health Officials Share How They Apply Evidence to Policymaking (NASHP 2017)
- The Center for Evidence-Based Policy
- The Value of Workforce Data in Shaping Nursing Workforce Policy: A Case Study From North Carolina (Fraher, Nursing Outlook, 2017)
- A series of briefs on using data to inform policy (UNC Sheps Center, 2017)
- The Rural Health Research Gateway’s toolkit on dissemination of rural health research (Schroeder 2015)
Q: How do you define and determine shortage?

A: A health workforce shortage means that there are not enough health care workers or not enough workers in specific professions, specialties, or settings to adequately serve patients' needs. Shortage is defined in different ways for different purposes. It is important to understand the difference between "shortage" and "maldistribution", particularly at the state and national level. Data and models may indicate that the nation or state has a sufficient supply of health professionals. However, this supply may not be evenly distributed across the country or state, creating pockets of shortage, especially in rural areas.

The Shortage Designation Branch at HRSA works with state Primary Care Offices (PCOs) to assign shortage designations to geographic areas, populations, and facilities that have too few providers and services; these are then eligible to receive certain federal resources. Designations include primary care, mental health, and dental Health Professional Shortage Areas (HPSAs) and Medically Underserved Areas and Populations (MUA/P). See https://bhw.hrsa.gov/shortage-designation/types for additional information on shortage areas, and http://www.healthworkforceta.org/resources/pco-resource-center for additional PCO resources.

Q: Is the number of health care jobs continuing to grow?

A: During the recession, healthcare jobs increased at the same time when many sectors were losing jobs. This trend is continuing to hold.

- The Bureau of Labor Statistics projects that employment of healthcare occupations will grow 18% from 2016 to 2020. This is faster than the average for all occupations and will add about 2.3 million new jobs.

- Altarum produces monthly Health Sector Economic Indicators briefs that monitor trends in health care employment, spending, and prices. Their latest employment brief shows continued growth in the number of health care jobs, with the greatest growth found in outpatient care centers.

- A November 2017 Health Affairs Blog discussed projected changes in health care employment under different policy proposals including the current law, H.R. 1628 American Health Care Act (AHCA), and ending cost-sharing reduction (CSR) payments to insurers. The projections and comparisons predict a loss of jobs under the AHCA and CSR payment reduction proposals. It concluded that continued monitoring is needed to ensure there are enough health care workers to meet the population's needs.

Q: I'm interested in allied health and administrative support workers. They're not always licensed. How do you count them?

A: For those professions, it may be necessary to conduct surveys, or rely on other data sources such as professional associations or the BLS, noting limitations as appropriate.

Q: What staff and resources are needed to undertake health workforce data collection and analysis?

A: This depends on many different factors, such as how many health professionals you want to track, the method used to collect data (licensure, survey, continuous monitoring, secondary data), the types of deliverables for which you're accountable, and organization structure. If the data system is embedded within a larger organization, such as a university or state government office, it is likely that some administration, finance, and infrastructure resources are already available for basic operation. If the data system is a stand-alone organization, you will need to secure funding.
In terms of staff, you may consider having a director to guide the work, make decisions, present results and acquire funding; one or more project managers/researchers to analyze data, write reports and present results; and a data manager to collect, clean and analyze data. Other positions may include communications specialist, visualization specialist, research assistant, administrative assistant, grants manager, and financial manager.

Additional resources needed include computer hardware and software for data management, statistical analysis, GIS, and graphic design.

Q: How do you fund health workforce data collection and analysis?

A: Data systems can be funded through state appropriations, private foundations, grants and contracts, and on a cost-recovery basis. Each funding mechanism has its challenges. State appropriations are tenuous; administrations and priorities change, and budgets get cut. Foundations are often geared to fund initiatives that show more tangible results. Grants are often time-limited. Cost-recovery is subject to demand for data and services, and limits the type of analyses and reports that you can do. Stakeholders who require data may be persuaded to fund the analysis costs to meet their specific needs, but they frequently are not willing or able to fund the fixed infrastructure costs. Consider the appropriate funding source for the specifics of your data collection effort, given the meaning and value of the project.

Citations


