Using Public Data for Regional Health Workforce Planning: A Toolkit

By Tim Bates, MPP, Lisel Blash, MPA, Elizabeth Mertz, PhD, MA

July 2025

Overview

Health workforce planning and development is a collaborative process requiring the engagement of multiple stakeholders with expertise across different domains, including health care systems, health professions training and education, and policymaking. This collaborative process also requires a common understanding of some key components of health workforce planning: current labor market conditions, the available supply of workers, characteristics of the population, and how changes in these components are likely to affect future conditions.

To support health workforce planning efforts, Policy at Healthforce Center at UCSF developed the following toolkit comprised of this technical report, a corresponding data dashboard, and webinar that describes the available health workforce data in California.

Purpose

The purpose of this toolkit is to help workforce planners access and understand the available data components for workforce planning: current labor market conditions, worker supply, and characteristics of the population. This toolkit is designed to support regional analysis (whether that be a geographic region, county, or metropolitan area) through databases, health workforce reports, and other publications.

Policy at Healthforce's Regional Health Workforce Planning Toolkit is comprised of the following three components:

1. **Health Workforce Data Dashboard**: An interactive data dashboard that allows users to search and access available workforce data by profession and geography.
2. **Health Workforce Data Planning Webinar**: An informational webinar for users interested in learning more about the available public data relevant to regional health workforce planning.
3. **Technical Report**: The technical report below provides comprehensive details of all available public data relevant to regional health workforce planning.

Available public data for regional health workforce planning

In the simplest terms, there are three main components to consider in the context of strategic health care workforce development: the current supply of health care workers and the educational pipeline that trains new entrants to the workforce, the current and future demand for health care workers, and the population that consumes health care services. Commonly referenced data points that describe the current health care workforce supply and pipeline include:

- Estimated employment and wages
- Number of actively licensed or certified professionals
- Education/training program completions
- Demographic characteristics of the current health care workforce

Demand for a health care workforce is driven by a range of factors – the population utilization patterns being the most important one. It is beyond the scope of this report to describe the various drivers of health care workforce demand and how those elements are modeled to produce forecasts of future demand. However, there are two readily available measures of labor demand that stakeholders can access:

- Current job postings
- Projected employment due to economic growth or retirements/attrition

With respect to the population of health care consumers, key data elements include:

- Demographic characteristics
- Socioeconomic status
- Health status/risk
- Expected population growth

This technical report gives an overview of public data sources associated with these different components and provides guidance on their respective strengths and limitations.

Labor market information and the Standard Occupation Classification (SOC) system

Key labor market data that are commonly used in strategic health care workforce planning include:

- Estimated occupational employment
- Estimated wages
- Current job openings
- Estimated employment by work setting
- Projected employment by occupation and work setting

This kind of labor market information is commonly organized according to the Standard Occupation Classification system (or SOC codes). The SOC system is the federal standard for classifying the workforce into occupational groups for the purpose of “collecting, calculating, or disseminating” data

about the workforce. The SOC system is hierarchical: there are more than 850 detailed occupational groups in the scheme that are organized into 23 broad categories. Almost all the occupations considered part of the health care workforce are included in three broad categories:

- 29-0000 Health Care Practitioners and Technical Occupations
- 31-0000 Health Care Support Occupations
- 21-0000 Community and Social Service Occupations

One important exception to this is *Clinical and Counseling Psychologists* (and *School Psychologists*); these occupations are included in 19-0000 Life, Physical, and Social Science Occupations.

The SOC system is updated periodically to incorporate new occupational roles or revise existing ones. However, there are some health care occupations that are not identified in the SOC system. For example, peer providers or care navigators or wellness coaches or doulas or licensed midwives (meaning not certified nurse midwives) have no SOC code. These occupations are captured somewhere in the labor market data, but they're rolled up into broader groups; they cannot be uniquely identified.

The educational pipeline and the Integrated Postsecondary Education Data System (IPEDS)

The Integrated Postsecondary Education Data System (IPEDS) is the most comprehensive source of information available to describe postsecondary education in the United States (US). IPEDS consists of a battery of surveys collecting data for a range of topics, but for health care workforce planning and development, the IPEDS Completions Survey is the most useful. This survey records the number of postsecondary degrees and certificates awarded by individual programs offered by postsecondary institutions. These data are a useful proxy for new entrants into the health care workforce. By law, every postsecondary education institution that participates in Title IV federal student aid programs is required to participate in IPEDS.

In the same way that labor market information is organized around the SOC system, IPEDS program completion data are organized using the Classification of Instructional Programs (CIP) system. Like the SOC system, the CIP system is a taxonomic scheme that organizes the content of different fields of study into uniform categories. Like the SOC system, CIP codes are organized hierarchically: the broad two-digit CIP codes encompass more detailed four-digit CIP codes, which in turn encompass even more detailed six-digit CIP codes. Most health professions education programs are identified within the following broad category:

- CIP 51 – Health Professions and Related Clinical Sciences

CIP 51 includes more than 200 different fields of study; however, it does not encompass all those that are commonly considered part of the health care workforce. Programs that train many of the behavioral health professions fields, such as clinical or counseling psychology, and clinical social work are found within the following broad categories:

- CIP 42 – Psychology
- CIP 44 – Public Administration and Social Service Professions

Despite the level of detail contained in the CIP system, in practice, the postsecondary education institutions reporting completions do not always utilize the six-digit codes to distinguish detailed fields of study within the programs they offer. This is an issue that often affects completions data for programs that train behavioral health professionals. For example, the following fields of study correspond with licensing categories used by the California Board of Behavioral Sciences:

- CIP 51.1503 – Clinical/Medical Social Work
- CIP 51.1505 – Marriage and Family Therapy/Counseling
- CIP 51.1508 – Mental Health Counseling

However, many postsecondary education institutions that offer master's degrees in these fields report program completions using a generic code for either clinical or counseling psychology, or social work, making it impossible to distinguish graduates who have completed training in these fields.

Nursing education provides further examples of this issue where IPEDS completions data fail to capture the full scope of the many different fields of study within nursing. More than twenty different fields of study related to nursing education are organized within the following broad category:

- CIP 51.38 – Registered Nursing, Nursing Administration, Nursing Research and Clinical Nursing

However, it is often the case that schools offering nursing education programs will report all nursing degrees/certificates awarded using the broad, generic CIP code:

- CIP 51.3801- Registered Nursing

At the master's level, this practice confounds efforts to describe fields of advanced practice nursing; there is no way to distinguish programs training nurse practitioners from those training clinical nurse specialists, nurse midwives, or nurse anesthetists. A related issue specific to nursing education is that IPEDS completions data do not distinguish between pre-license baccalaureate programs and post-license RN (registered nurse) to BSN (Bachelor of Science in Nursing) programs; both are simply reported as bachelor's degrees awarded. This lack of detail makes it impossible to determine which BSN program graduates are new to the RN workforce.

For some fields of study, training providers may be outside of the traditional postsecondary education system. For example, the direct care workforce is largely trained by adult vocational schools, community-based organizations, or through on-the-job training programs delivered by employers. This is also true for community health workers (CHWs); very few of the programs that train CHWs report student data through IPEDS. In certain fields of allied health, education and training is predominantly offered by private for-profit institutions, many of which do not participate in Title IV student aid programs; these providers will not show up in IPEDS. It is important to understand that for some segments of the health care workforce, IPEDS data may not capture the level of detail needed to fully describe the health workforce educational pipeline.

Where possible, it is helpful to compare the set of postsecondary institutions in IPEDS reporting program completions in a specific field of study with a list of all known programs (e.g., a list of approved schools maintained by the licensing board that regulates that profession, a list of accredited programs maintained by the agency that accredits programs of study in that field). This practice will provide a

sense of how completely (or incompletely) the IPEDS completions data describe the educational pipeline.

Aggregated labor market and educational pipeline data

Data aggregation sites pull together information from various sources into a single report and are a useful way to obtain benchmark statistics that describe both characteristics of the labor market and educational pipeline. The Centers of Excellence for Labor Market Research (COE) are regional, technical assistance centers that support the California Community Colleges (CCC) and workforce development professionals by producing occupation-specific (SOC-based) reports and tools that provide current, regional labor market and educational pipeline (programs identified by CIP code) information designed to help colleges with strategic planning and program development. There are nine different regional centers that cover the entire state of California.

The COE reports and other analytic tools are meant to support California's community colleges, which means the emphasis is on health care occupations that are trained within the community college system. As the regional centers are primarily responsive to the strategic needs of the region's colleges, they have not produced reports for every single health care occupation. However, if there is an available report, that is a strong indicator that the occupation in question has been identified as a regional priority by stakeholders. (The COE are also available to engage in contract work on behalf of entities outside of the community college system.)

In addition to the occupation-specific reports, the COE have two other analytic tools that can be used to understand regional labor market conditions: one focused on projected demand for employment, the other on education program completions. Both are Excel-based tools that allow users to filter results by region, county, and either SOC (projected demand tool) or CIP (program completions tool) code. The projected employment demand tool is populated with data sourced from the California Employment Development Department, Labor Market Information; the educational pipeline tool is populated with data sourced from the CCC and IPEDS. Both tools can be downloaded, and each comes with a comprehensive user guide.

Another useful site for finding aggregated information is the O*NET OnLine tool. It overlaps with the kind of information provided by the COE but does so for the full scope of occupations specified in the SOC system. Performing a search for a given occupation produces information on a range of topics, including total employment, wages, job openings (and the employers posting job openings), educational programs, and program completions. A helpful feature of the O*NET tool is the ability to use a zip code to filter results for local data.

California Department of Health Care Access and Information

The California Department of Health Care Access and Information (HCAI) has a broad set of priorities related to health care workforce and health professions education. As a result of legislation in the last several years, however, HCAI now has a much stronger emphasis on health care workforce data collection for licensed and certified health professionals in California, producing workforce analyses and providing guidance on a range of health care workforce and education policy issues. The HCAI website hosts several new health care workforce data dashboards organized around three main topics: workforce racial and ethnic composition, languages spoken, and educational pathways. Within each of

these topical areas, users can manipulate by selecting different parameters (e.g., license type, license issue date, geographic region) to produce output describing licensed or certified health care professionals in California.

HCAI also produces an annual report summarizing the wide range of information it collects describing licensed and certified health professions. (See Appendix A for a full list of the licensed and certified professions). It is beyond the scope of this report to detail the full range of data elements collected by HCAI, but those that are central to health care workforce planning include:

- Characteristics of the education pipeline
- Employment status and how they spend their time at work
- Retirement estimates
- Age distribution
- Racial and ethnic composition
- Languages spoken

Included in the appendices of HCAI's annual report are regional data, and a particularly useful table that HCAI staff have produced is one that shows a regional *distribution index value* for each licensed or certified profession. This index measures the magnitude of the difference in a region's share of licensed providers relative to its share of the general population. For example, a distribution index value of 1 indicates equality (the region's share of a given licensed provider is proportional to its share of California's general population); index values below 1 indicate a low supply of licensed providers relative to population, while index values above 1 indicate high supply of licensed providers relative to population. These index values provide information about where there are likely imbalances in provider supply versus demand.

Other HCAI data sets that may be useful for the purposes of workforce planning include:

- Registered Nurse Shortage Areas in California – an Excel file that identifies and ranks designated RN shortage areas across the state
- Health Professional Shortage Areas in California – an Excel files that identifies and ranks designated primary care (physicians, nurse practitioners, physician assistants), mental health, and dental health professional shortage areas

Although HCAI is an important source of data describing California's health care workforce, there are several occupations regulated by agencies within the California Department of Public Health (CDPH) that are not yet represented in the HCAI data. These include:

- Certified nursing assistants
- Home health aides
- Clinical laboratory science professionals
- Radiologic/diagnostic imaging professionals

The CDHP licensing and certification department maintains a list of approved training providers for certified nursing assistants and home health aides, so it is possible to identify a local training pipeline.

As well, Laboratory Field Services (an agency within CDPH) maintains a list of approved training providers, and for a fee will produce a census of current licensees. Finally, there are health professions that do not require licensure or certification, and so they are not identified in data sourced from either HCAI or CDPH (though labor market and education data derived from other sources may be available). Examples of unlicensed health care professionals include:

- Diagnostic medical sonographers
- CHWs
- Doulas

Other HCAI activities

As part of its growing emphasis on strategic workforce development, HCAI is also engaged in building analytic tools to help inform health care workforce and education policymaking. Current efforts are focused on modeling the supply and demand for the behavioral health care workforce, and the registered nursing workforce (and educational pipeline). These models are designed to inform policy making at the local (county) level. Links to slide deck presentations describing this work and initial findings from public meetings can be found in Appendix B. Related to this work is the recent formation of the California Health Workforce Education and Training Council, established as part of the Song-Brown Healthcare Workforce Training Program. The council is a group of 18 members who are charged with "helping coordinate California's health workforce education and training to develop a health workforce that meets California's health care needs." The members are appointed by the governor, the state assembly speaker, the state senate rules committee, and other state government agencies. The Council meets on a regular basis to discuss HCAI program activities and these public meetings typically feature invited speakers who have expertise on key health workforce and education policy issues; they are an effective way to stay informed about HCAI priorities and progress toward its goals.

Other sources of health care workforce data

This section highlights additional sources of data frequently used by health services researchers to describe California's health care workforce and educational pipeline.

The California Board of Registered Nursing School Survey Interactive Dashboard is the most complete source of data describing pre-licensure nursing education in California. The dashboard provides data covering a ten-year period, and the data are available at the regional level. It includes more than 30 different data elements covering a range of topics, including:

- Student enrollments
- Applications for admission
- Program completions
- National Council Licensure Examination (NCLEX) pass rates
- Student and faculty demographics

The California Board of Registered Post-Licensure Nursing Program Data Summary and Historical Trend Analysis is the companion data to the pre-licensure nursing school interactive dashboard, available as a static report. It provides a comprehensive look at student, faculty, and program characteristics for post-licensure nursing programs in California.

The California Board of Registered Nursing Survey of Registered Nurses has been fielded continuously since 1990. It is the most comprehensive source of information describing the state's RN workforce. Most of the data are presented at the state level, but it does provide some regional information.

The Health Services and Resources Administration maintains data dashboards that allow users to identify sub-state geographic areas where the population is underserved by primary care providers, or have been designated as a shortage area for primary care physicians, dentists, or mental health services providers, or dashboards that you can use to visualize the retention and distribution of alumni of Bureau of Health Workforce programs (e.g., National Health Service Corps).

The Bureau for Private Postsecondary Education (BPPVE) is an agency within the California Department of Consumer Affairs that is responsible for regulating the private postsecondary institutions operating in California. Schools are required to file an annual report with program-specific data, including the number of students who began the program and the number who graduated, as well as information about the number of graduates employed in a field related to their program of study. In cases where a private postsecondary institution does not participate in federal student aid programs and therefore is not identified in IPEDS data, the BPPVE reports may be an alternative source of information to help describe the educational pipeline into the local health care workforce.

Population data

Population differences in health status/risk, socioeconomic status, gender, race/ethnicity, and age influence patterns of health care services consumption, thus demographics are a key driver of demand for a health care workforce.

US Census data

A useful place to start understanding the demographic profile of the local population is the US Census Quick Facts website, which uses data sourced from the American Community Survey (ACS). The Quick Facts tool gives users access to commonly referenced population characteristics, including racial/ethnic composition, as well as socioeconomic, disability, and health insurance status. The Quick Facts tool allows users to produce statistics for the smallest geographical unit for which data are available. However, it is important to keep in mind that, depending on the size of the geography, the data output may be based on a five-year average, as opposed to the most recent single year of data. Small geographic areas rely on a sample that aggregates five-years of ACS data to produce reliable estimates and avoid disclosing confidential information; the output tables will indicate the timeframe of the estimate.

For users who want to access the full scope of data collected by the Census Bureau, there is the Explore Census Data tool. It is a more complex interface and may take some practice to navigate and produce the desired output. It is recommended that users explore the video tutorials and the FAQs to learn about best practices for using the site.

California Health Interview Survey

The California Health Interview Survey (CHIS) is a large-scale household survey conducted in six different languages covering more than 100 different population health topics, including access to care, health and dental insurance coverage, and health conditions and behaviors. The data collected by CHIS are publicly available, though a registered account is required. Users can access CHIS data through its AskCHIS dashboard and build custom data queries, which can be structured to produce output at the county or regional level (and include reference geographic areas).

CDC Places: Local Data for Better Health

The CDC Places database is an interactive web tool that provides population health status and related information for geographic areas as small as the zip code. It is recommended that users become familiar with how to use the data portal before attempting to access and create customized output.

California Department of Finance population projections

Changes in population size and demographics are an important driver of demand for health care services and thus the health care workforce. The California Department of Finance Demographic Research Unit publishes population projections by age, sex, and by race and ethnicity for all California counties, available as downloadable Excel files.

Data limitations

Some limitations of using publicly available data have been acknowledged earlier in this report, but it is useful to recap them and acknowledge other common challenges associated with using public data to conduct health care workforce analysis.

- There is very limited information available for some occupations; peer providers, doulas, CHWs, licensed midwives, and the broader public health workforce are all examples. If the occupation is not uniquely identified by an SOC code, the Bureau of Labor Statistics is not collecting any labor market information for it. Similarly, if the occupation is not licensed or certified by the state, there will be very limited information available through any source.
- SOC codes do not perfectly correlate with licensing/certification categories. For example, labor market data describing behavioral health occupations often include unlicensed personnel.
- If the educational pipeline is made up of training programs that are either outside of the postsecondary education system, or largely for-profit private institutions that do not participate in Title IV student aid programs, student data is difficult to obtain (not available via IPEDS).
- In general, state and national level data are typically of higher quality and easier to obtain; you may not be able to find the data you want for the level of geographic detail that you are focused on (or the data may lag for the current period).
- Occupation-specific retention and vacancy data are not readily available; there are hospital and clinic systems that collect this information, but it is typically proprietary information. There are many such sources of high-quality data (e.g., American Association of Medical Colleges) that are available for a fee.

About us

Policy at Healthforce promotes health workforce diversity and economic opportunities in California through a responsive, community-informed research and policy agenda rooted in social justice, with support from The California Endowment. Policy at Healthforce is part of Healthforce Center at UCSF, a trusted partner to funders, policymakers, and health care organizations, delivering impactful research, evaluation, policy insights, and capacity building programs. Grounded in equity and built on deep relationships across California's health care landscape, our work breaks down silos and drives system transformation—advancing better health for all.

Find more research and evidence from Healthforce Center on the health workforce.

Appendix A: HCAI occupational licensing/certification categories

Allied health

- Audiologist
- Chiropractor
- Acupuncturist
- Podiatrist
- Optometrist
- Occupational therapist/assistant
- Pharmacist/pharmacy technician
- Physical therapist/assistant
- Polysomnographic technologist/technician
- Respiratory care therapist
- Speech language pathologist (SLP)/SLP assistant

Behavioral health

- Licensed clinical social worker
- Associate clinical social worker
- Licensed marriage & family therapist
- Associate marriage & family therapist
- Licensed professional clinical counselor
- Associate professional clinical counselor
- Psychologist
- Registered psychological associate
- Licensed educational psychologist
- Psychiatric mental health nurse
- Psychiatric technician

Medicine

- Naturopathic doctor
- Osteopathic physician & surgeon
- Physician & surgeon
- Physician assistant

Nursing

- Public health nurse
- Registered nurse
- Licensed vocational nurse

Advanced practice nursing

- Clinical nurse specialist
- Nurse anesthetist
- Nurse midwife
- Nurse practitioner
- Licensed midwife

Oral health

- Dentist
- Orthodontic assistant
- Registered dental assistant
- Registered dental assistant in extended function
- Registered dental hygienist
- Registered dental hygienist in extended function
- Registered dental hygienist in alternative practice

Appendix B: Links to HCAI supply/demand modeling presentations

Overview of the modeling initiative

https://hcai.ca.gov/wp-content/uploads/2024/03/Agenda-Item-5-HCAI-Strategy.pdf

Modeling the behavioral health workforce

https://hcai.ca.gov/wp-content/uploads/2024/10/Agenda-Item-5-Summary-of-Supply-and-Demand-Model-and-Detailed-Review-of-Behavioral-Health-Workforce-Strategy.pdf

Modeling the registered nursing workforce and nursing education pathways

https://hcai.ca.gov/wp-content/uploads/2024/11/Agenda-Item-8a-Nursing_Modeling-updates-and-county-level-findings_Nov_2024_Council_Mtg.pdf

https://hcai.ca.gov/wp-content/uploads/2024/11/Agenda-Item-8b-Nursing_Funding-and-application-of-findings_Nov_2024_Council_Mtg-1.pdf

https://hcai.ca.gov/wp-content/uploads/2024/11/Agenda-Item-8c-Nursing_Results-of-ADN-Expansion-Project_Nov_2024_Council_Mtg.pdf