CALIFORNIA
HEALTHCARE
FOUNDATION

Improving Health Data Access:

State Policymakers Weigh In

April 2013

Improving Health Data Access:
State Policymakers Weigh In

Prepared for
California HealthCare Foundation

by
Janet M. Coffman, PhD
Margaret Fix, MPH
University of California, San Francisco

April 2013

About the Authors

Janet M. Coffman, MA, MPP, PhD, is associate professor at the Philip R. Lee Institute for Health Policy Studies and the Department of Family and Community Medicine at the University of California, San Francisco.

Margaret Fix, MPH, is research associate at the Philip R. Lee Institute for Health Policy Studies at the University of California, San Francisco.

Acknowledgments

The authors would like to thank Claire Brindis, DrPH, director, Philip R. Lee Institute for Health Policy Studies, for sharing her expertise in needs assessment, and Claire Will, PhD, for providing administrative support for the project.

This project also builds on work funded by the University of California, San Francisco Clinical and Translational Science Institute, part of the Clinical and Translational Science Award program funded by the National Center for Advancing Translational Sciences (Grant Number UL1 TR000004) at the National Institutes of Health.

About the Foundation

The **California HealthCare Foundation** works as a catalyst to fulfill the promise of better health care for all Californians. We support ideas and innovations that improve quality, increase efficiency, and lower the costs of care. For more information, visit us online at www.chcf.org.

Contents

I. Introduction

Beginning with the National Oceanic and Atmospheric Administration's free-of-charge release of weather data more than 30 years ago, federal agencies have been publicly releasing selected data on a wide range of topics. In 2009, the White House launched an Open Government Directive to increase the speed and scope of such public releases. In response to this directive, the United States Department of Health and Human Services (HHS) inaugurated its Health Data Initiative in 2010. This initiative releases data to the public in a number of ways and engages software developers in creating new methods for disseminating data to consumers and other stakeholders.

The California HealthCare Foundation (CHCF) is supporting national efforts to increase the availability of health care data and has started its own initiative to improve access to data in California. CHCF's Free the Data initiative, launched in December 2011, is especially concerned with improving access to data for policymaking, which becomes particularly critical as California makes decisions regarding implementation of the Affordable Care Act (ACA).

To obtain input from state-level California policymakers on their experiences with state health care data, CHCF engaged the Philip R. Lee Institute for Health Policy Studies at the University of California, San Francisco, to conduct a needs assessment. The project's objectives were to delineate state-level policymakers' priorities for improving access to data and to identify barriers to and facilitators of expanded data access. This report describes the research team's methods for eliciting input from California policymakers and summarizes its findings regarding data currently available from state agencies, policymakers' use of that data and their priorities for improving data access, and barriers to expanded access. The paper also includes a list (Appendix B) of current sources of machine-readable California health care data.

CHCF's Free the Data Project

Launched in 2011, CHCF's Free the Data initiative seeks to ensure that more health care data are publicly reported in California while helping to catalyze development of tools to access, analyze, and communicate these data. The initiative works across a number of fronts to achieve progress on many of the issues discussed in this report. This work includes:

- Identifying best practices and lessons from other efforts to provide access to government data
- Demonstrating innovative approaches to data visualization and display
- Developing an inventory of state and local health care data and identifying obstacles to access
- Supporting efforts to integrate health care data into websites and mobile apps for consumers, providers, payers, and other interested audiences
- Encouraging broader use of health care data by the media
- Aiding health policy audiences with tools to use data in their work
- Providing consumers with better access to data about health care quality

To learn more about the initiative, visit the Free the Data pages of CHCF's website, www.chcf.org.

II. Methodology

MOST OF THE INFORMATION FOR THE assessment in this project was obtained through interviews with producers and users of health care data in California. In-person and telephone interviews were conducted between February and May 2012 with staff from the California State Legislature, legislative support organizations (e.g., Legislative Analyst's Office), executive branch agencies, and consumer advocacy organizations. (A complete list of interviewees appears in Appendix A.) Discussion guides were used during the interviews to ensure that all pertinent topics were addressed, but the order of topics varied with the flow of conversation.

The interview questions focused on data pertinent to California policies regarding access, cost, and quality of health care services. The topics of the questions for representatives of state agencies and organizations that *produce* health care data included:

- Subjects on which data are collected
- Formats in which data are released (e.g., datasets, query tools, reports)
- Capacity of agencies to analyze their data
- Plans to expand access to data (if any)
- Barriers and facilitators to expanding access to data

The topics of questions for interviewees who are representatives of statewide agencies and organizations that *use* health care data to inform policy decisions included:

- Types of data-related products that organizations generate
- Sources of data used
- Formats in which data are accessed
- Satisfaction with the availability of data

The research team also reviewed the websites of state government agencies that have jurisdiction over health care. This review focused on identifying sources of machine-readable health care data, such as those in Excel and SAS files, because machine-readable data can be more easily utilized across a wide range of platforms than data available only through PDF files or other static formats. In addition, interactive tools can be created to enable users to query machine-readable data to answer questions more quickly, to visualize trends over time or geographically, and to tailor analyses to specific interests. (A descriptive list of California sources of machine-readable health care data appears in Appendix B.)

In June 2012, the research team presented findings from the interviews to a convening of key stakeholders. The presentation was followed by a discussion about participants' priorities for investment in improving access to health care data for policymaking, including better access to existing data and collection of additional data. Participants' input has been incorporated into this paper's findings.

III. Findings

Data Currently Available from State Agencies

The topics on which California state government agencies currently provide health care data can be grouped into seven major areas:

- Participation in Medi-Cal, Healthy Families, and other public programs
- Health plan enrollment and finances
- Use of health care services
- Health care workforce
- Quality of care
- Health status
- Vital statistics

(The sets of data that are offered in machine-readable format, and their sources, are listed in Appendix B.)

State agencies release data in a wide range of formats, including tables and pivot tables, query tools, and datasets. Subject to their resources, some agencies also provide customized analyses of their data. For most agencies, static tables in Excel or PDF format are the most common modality in which data are released. While static tables can be helpful if they address matters of interest to policymakers, in general they cannot answer as wide a range of questions as interactive query tools.

Agencies provide two major types of query tools that can give policymakers some flexibility to customize their analyses without having statistical software skills. For example, Excel pivot tables enable users to extract descriptive data for specific populations, geographic areas, and types of health care providers. Also, online point-and-click interfaces provide a user-friendly format that facilitates comparisons across groups.

Some agencies also provide datasets that can be analyzed using statistical software packages. Datasets may be de-identified, limited, or confidential and, in some cases, most notably with hospital discharge data from the Office of Statewide Health Planning and Development (OSHPD), both de-identified and confidential versions of the same dataset may be available. De-identification is an important strategy for making data available while simultaneously protecting privacy. However, de-identified datasets may not contain all the information needed to answer some important policy questions. For example, de-identified OSHPD datasets cannot be used to analyze hospital readmissions because they do not distinguish between initial admissions and readmissions of the same patient. Thus, they cannot be used to compare rates of readmission across hospitals or to assess how hospitals are responding to incentives to reduce preventable readmissions.

Current Data Use by Policymakers

Interviewees reported that they use data from a wide range of sources to inform state-level decisions about health care policy. These sources include 11 state government agencies, as well as state licensing boards for health professionals. For some topics, interviewees also examine data from national sources, such as the United States Census Bureau, the Centers for Medicare and Medicaid Services, and the National Conference of State Legislatures. Interviewees also use data from philanthropic foundations, most notably CHCF and the Kaiser Family Foundation. (A complete list of health care data sources used by interviewees appears in Appendix C.)

Interviewees use health care data to prepare multiple types of work products. The specific work products vary across the types of organizations engaged in policymaking. For example, members of the staffs of state legislative committees primarily produce analyses of bills that their committees consider. In contrast, legislative support organizations and advocacy groups produce multiple types of work products, including memos, fact sheets, issue briefs, and reports, with some of the data displayed in tables and graphs.

Most interviewees prefer to receive data in the form of reports or query tools, rather than in a more raw format, because their organizations do not perform their own statistical analyses of datasets. Some organizations lack staff with training in data analysis. Others need to make decisions too quickly to program datasets. Some prefer reports from which they can extract key statistics without doing any of their own analyses. Others like online query interfaces, such as AskCHIS, the query tool for the California Health Interview Survey. This is an example of a well-designed interactive tool that enable users without expertise in programming statistical software to generate their own queries.

Priorities for Improving Access to Health Care Data

Interviewees offered numerous suggestions for making health care data more available to policymakers. Some of these suggestions concern improving access to existing data sources while others involve collecting types of data currently unavailable. Interviewees suggest that efforts in both spheres are important to policymakers and that their relative value varies with the extent to which existing data sources provide pertinent information.

Access to Existing Data

Some of the interviewees' suggestions for improving access to existing health care data were general, while others focused on specific topics.

General. Multiple interviewees recommended that state agencies make more health care data available electronically in machine-readable formats. Although some interviewees prefer to use static tables and reports, others would like to have interactive query tools that would enable them to use data to answer their own questions. Interviewees repeatedly cited AskCHIS. Some interviewees also recommended expanding availability of mapping capabilities because of the way maps can help policymakers visualize data.

In addition, some interviewees recommended increasing the sharing of health care data across agencies and between state agencies and city and county governments. A desire for the creation of a single, "one-stop shopping" portal for obtaining health care data collected by all state agencies also was noted. Such a portal has been created by the federal government with its Healthdata.gov, which is a clearinghouse for data from such sources as the Centers for Medicare & Medicaid Services, Centers for Disease Control and Prevention, Food and Drug Administration, and National Institutes of Health (www.healthdata.gov). Those interviewees believe that if policymakers only had to go to one web portal, they might use more data to inform policy decisions.

Medi-Cal. Many interviewees, including staff of the Department of Health Care Services (DHCS), would like to see expanded access to Medi-Cal data. DHCS currently provides only a limited number of reports and pivot tables on its website. Policymakers whose questions go beyond the scope of these resources must ask DHCS staff to generate custom analyses for them. Obtaining timely responses to such requests can be challenging due to competing demands faced by DHCS's data analysis staff and the complexity of Medi-Cal data. Also, DHCS's current contract with the vendor that maintains its datasets limits DHCS staff access to its own data. In particular, some interviewees would like DHCS to develop a dashboard summarizing key performance measures. Others recommended creating a query tool, similar to AskCHIS, that would enable them to answer a wider range of questions about Medi-Cal. Topics of great interest to interviewees include enrollment, provider networks, use of health care services, and health status, with some interviewees wanting the ability to access beneficiary-level data on health status and utilization.

Other public programs. Several interviewees are interested in improving access to existing data on public programs other than Medi-Cal. One example is the impact of the 2011 realignment of correctional services under which responsibility for custody, treatment, and supervision of persons convicted of nonviolent, nonserious, nonsex crimes has been shifted from state to county correctional systems. Several interviewees expressed interest in the impact of this realignment on county budgets, particularly on the availability of resources to provide medical and mental health services.

Health plans. Interviewees would like to have better access to information that the state's Department of Managed Health Care (DMHC) and the California Department of Insurance (CDI) collect on health plans. They are especially interested in more detailed data on enrollment. DMHC has an interactive query tool on its website that can be used to retrieve financial and enrollment data for the health plans it regulates. However, the usefulness of this query tool is limited because it retrieves only statewide data and generates Excel tables that users must then manipulate to display and analyze results. Also, CDI does not offer a tool for obtaining enrollment data.

Some interviewees expressed interest in obtaining access to data regarding health care services provided to state and local government employees and their dependents who have health insurance through the California Employees Public Retirement System (CalPERS). Analysis of data on health care utilization by people covered through CalPERS could provide insights into patterns of care delivery in private health plans, because CalPERS offers a choice of several health maintenance organization and preferred provider organization health plans. These data would also provide policymakers with information to help monitor health expenditures for state and local employees and their dependents.

Quality of care. Many interviewees would like better access to data on the quality of care that patients receive. One challenge for policymakers and other stakeholders interested in quality of care is that different data are collected by different state agencies. For example, OSHPD reports data on the Agency for Healthcare Research and Quality's Quality Indicators extracted from hospital discharge records, whereas the California Department of Public Health (CDPH) collects and reports data on hospital-acquired infections, and the Office of the Patient Advocate reports data on Healthcare Effectiveness Data and Information Set (HEDIS) measures of quality of care at the health plan level. Some interviewees recommended creation of a single web-based portal through which users could access data on quality of care from multiple state agencies. They believe that policymakers and other stakeholders would be more likely to use data on quality of care if they only had to go to one place to find them. Some interviewees were especially interested in the public reporting of machine-readable data on hospital-acquired infections. Interviewees also recommended linking data on cost and quality of care collected by various state agencies.

Collection of New Data

Interviewees had a number of suggestions regarding health care data that state agencies do not currently collect or make available.

Medi-Cal. Interviewees are especially interested in collection of additional data on Medi-Cal. Specific topics mentioned include the transition of aged and disabled Medi-Cal beneficiaries to managed care, the transition of Healthy Families enrollees to Medi-Cal, and the implementation of the Low-Income Health Programs. Interviewees would also like to obtain data on take-up rates, demographic characteristics, and use of health care services among people who will become newly eligible for Medi-Cal under the ACA.

Health plans. Additional suggestions from interviewees concerned data on health plans, interest in which is largely driven by the ACA. Interviewees are especially interested in health plan data from the California Health Benefit Exchange because the exchange, as a new entity, will develop its data collection and dissemination infrastructure from scratch. Some are particularly interested in data on the use of health care by people enrolled in health plans offered through the exchange. The reasoning here is that these data could help policymakers assess whether risk adjustment mechanisms are adequate to prevent adverse selection. Others would like to monitor the affordability of health insurance purchased through the exchange and other sources.

Some interviewees recommended creation of an all-payer database that would contain claims and encounter data for all people with health insurance. Currently, California does not collect comprehensive claims data for privately insured persons, and data on claims paid for people with publicly supported health insurance (e.g., CalPERS, Medi-Cal) are not readily available or reported in a standardized manner. Some interviewees expressed interest in obtaining data on the actual prices that health plans pay providers for health care services. Others would like the state to collect data on the design of health insurance benefits, such as cost-sharing requirements (e.g., deductibles and copayments) and the types of health care services health plans cover, so that policymakers could compare the effects of variation in benefit design on use of health care services.

Health care workforce and facilities. Some interviewees recommended expanding the types of health care facilities about which state agencies collect data. Certain interviewees expressed interest in expanding OSHPD's authority to collect and disseminate data on ambulatory care to encompass types of providers not currently reported, such as physician offices; OSHPD has authority to collect data only from emergency departments and licensed clinics. Others called for standardizing and expanding the collection of data on the health care workforce to encompass demographic characteristics, hours worked, practice location, and participation in Medi-Cal and other public programs. Monitoring participation in Medi-Cal is especially important to some interviewees because the ACA will substantially increase the number of Californians enrolled in Medi-Cal. At present, only two-thirds of California physicians treat any Medi-Cal patients, raising questions about whether the supply of physicians accepting Medi-Cal patients will be adequate to meet the needs of new enrollees.

Quality of care. Interviewees expressed strong interest in collecting new data on the quality of care provided to Californians. Some interviewees advocated for the development of metrics to monitor the quality of care provided by health plans participating in the California Health Benefit Exchange. Such metrics could be used by the legislature and the exchange to monitor quality of care and could also be used by the exchange to inform contract specifications and negotiations. Several interviewees indicated that focusing on the exchange could be especially productive because its leadership is interested in quality and because it is a new organization still designing its data systems.

Other interviewees expressed interest in adding to the types of quality of care data that state agencies currently collect. Several suggested expanding the range of indicators used to assess the quality of care provided by Medi-Cal managed care plans because they believe the HEDIS indicators are insufficient. Others recommended expanding OSHPD's efforts to disseminate data on the quality of hospital care to include physician-level data because there is evidence that quality can vary substantially across physicians within individual hospitals.

Barriers to Expanding Access to Health Care Data

Interviewees described a number of barriers to expanding collection of and access to health care data generated by state government agencies.

Legal Restrictions

In some cases, state laws, regulations, or judicially protected proprietary or privacy rights prohibit or limit the extent to which state agencies can collect or release health care data. For example, OSHPD has statutory authority to collect and release data on hospitals' charges for services but not data on the prices that health plans pay for services. Restrictions on release of data are often more stringent when data from different sources are matched with one another. This is due to concerns that combining data from multiple sources increases the risk that the data could be used to identify individuals.

Interpretation of state law by attorneys in state government agencies often determines whether health care data can be released and the level of detail that may be provided. When doing so, they weigh the risks and benefits associated with releasing data but their assessments do not always coincide with those of staff who oversee data collection and dissemination efforts. Agency program staffers sometimes feel that the attorneys unnecessarily constrain staff efforts to provide policy-relevant data to stakeholders. Further, attorneys' judgments regarding similar data are not always consistent across agencies.

Limited Resources

In recent years, the state budget crisis has made it more difficult for state agencies to devote resources to the collection and dissemination of health care data. Some agencies do not have sufficient staff with training in data analysis to disseminate data in a timely fashion; others have outdated information technology systems that limit staff's ability to analyze data and to make it available in user-friendly formats. Interviewees suggested that data collection and dissemination efforts are especially vulnerable in agencies in which data sharing is not considered a primary mission; when faced with budget constraints, these agencies tend to allocate fewer resources to data collection and dissemination relative to other functions.

Data Source Limitations

Some agencies have difficulty accessing and analyzing data necessary to fulfill their missions. Data on some topics are reported and stored on paper; for example, CDPH's data on licensed health care facilities is available only in hard copy. In addition, some data are not reported in a standardized manner, which makes it difficult to compare data across counties or institutions. Other data are not reported in formats that are easily understood by consumers or other stakeholders, and agency staff do not always have sufficient time or resources to convert such data into comprehensible information.

Lack of Consistency Within and Across Agencies

Information technology systems vary widely across state agencies that collect health care data, and these systems are not always compatible. Also, agencies sometimes use different categories for classifying data on the same variables, such as race/ethnicity and other demographic characteristics. In addition, some departments, most notably CDPH, rely heavily on categorical funding (grants and other sources of funding that can be used only for certain purposes). In these departments, staff members have difficulty collecting and aggregating data in standardized formats because certain categorical funds are intended to be used only to address specific topics and may also limit the manner in which the data are displayed.

Lack of Standardized Unique Identifiers

The unique identifiers used to identify individual patients, providers, or facilities are not consistent across state agencies, which makes it difficult to link their datasets. Even where standardized unique identifiers exist, such as social security numbers, agencies may be reluctant to use them due to privacy concerns.

Lack of Awareness Regarding Need and Opportunities for Collaboration

Some agencies seem to be unaware of opportunities or requirements for cross-agency collaboration regarding dissemination of health care data. OSHPD's Healthcare Workforce Clearinghouse provides a good illustration of this challenge. OSHPD staff responsible for the clearinghouse report that they had to devote extensive time to educating their colleagues at licensing boards and other state agencies about the statute that requires OSHPD to obtain data on supply, demand, distribution, demographic characteristics, and training of health professionals from agencies that collect such data.

Reluctance by Stakeholders

Lack of consensus among key stakeholders can limit the collection and release of health care data. Key stakeholders may resist reporting to state government agencies data that they consider proprietary, especially if the data could reflect badly on the stakeholders. Compromises with stakeholders over such contested data may lead agencies to publicly report data that does not fully address policymakers' questions. For example, state law requires hospitals to report data to OSHPD only on charges; they are not required to report data on prices paid by health plans, which would be helpful for assessing variation in the cost of hospital care. Similarly, OSHPD has authority to collect and report data on the quality of care provided by hospitals but not the quality of care provided by individual physicians practicing in hospitals. Such data could help policymakers determine whether hospitals' aggregate scores on quality measures mask substantial variation in quality across physicians.

Concerns About Misuse of Data

Concerns about misuse of data have made some state agencies reluctant to release data in machine-readable formats. For example, staff of DHCS asserted that financial data regarding Medi-Cal are so complex that people who do not have high technical familiarity with the program could inadvertently reach inaccurate conclusions. Interviewees also reported that CDPH has cited concerns about misinterpretation of data as a reason for not releasing machine-readable data on health care–associated infections.

Insufficient Demand for Data

Several interviewees stated that improving access to health care data is insufficient to increase the use of data to inform health care policy decisions unless efforts are also undertaken to create demand for such data. These interviewees observed that policymakers' demand for data varies widely. Some policymakers are very adept at seeking pertinent data, whereas others are unaware that data exist that could inform decisions. Knowledge of data sources and data analysis methods also varies widely. To increase demand for data, these interviewees suggested that greater efforts are needed to educate policymakers, especially legislative staff, about useful sources of data. In this regard, they also recommended providing concrete examples of innovations in data collection, analysis, and display that can help policymakers use pertinent data, such as dashboards for key indicators and policy simulation models for conducting scenario analyses.

IV. Moving Forward

BASED ON THIS PROJECT'S FINDINGS, three areas can be identified for concerted efforts to improve collection of and access to state health care data:

- **Provision of resources to improve access to health care data.** Key topics include data from Medi-Cal, and from DMHC and CDI on health plan enrollment, benefit design, and other topics. Access can be improved by ensuring that all data are stored in machine-readable format, by standardizing data collection across state health departments, and by building query tools and dashboards that would make it easier for government agencies to publish public domain data and for policymakers to find the data they need. Alongside improving access to existing data, there is considerable interest in collecting new data, particularly data related to implementation of the Affordable Care Act in California and physician-level data on quality of care.

- **Advocacy for legal changes to facilitate data collection and dissemination.** A matrix of state and federal law limits access to some data that can be pertinent to policymaking. Maximizing release of data within these legal restraints may require convening privacy officers, attorneys, and program managers from state agencies to develop a consensus on interpretation of the laws and regulations that limit and authorize the release of health care data.

- **Communication among stakeholders regarding innovations in data collection and dissemination.** A health data network could be established in California to share ideas and lessons learned about expanding data access. Similarly, a series of convenings and webinars could be launched to review best practices and to showcase the work of other states.

As part of its recently launched Free the Data initiative, the California HealthCare Foundation is funding projects related to many of the aforementioned goals. Visit www.chcf.org to obtain regular updates.

Appendix A: Project Interviewees

Brent Barnhart, JD
Director
California Department of Managed Health Care

Teri Boughton
Chief Consultant, Health Committee
California State Assembly

Janette Casillas
Executive Director
California Managed Risk Medical Insurance Board

Toby Ewing, PhD
Consultant, Governance and Finance Committee
California State Senate

Len Finocchio, DrPH
Associate Director
California Department of Health Care Services

Kim Flores
Policy Consultant, Office of Research
California State Senate

Scott Graves, PhD
Senior Policy Analyst
California Budget Project

Betsy Imholz, JD
Director, Special Projects
Consumers Union

Peter Lee, JD
Executive Director
California Health Benefit Exchange

Barbara Marquez
Deputy Director
California Office of the Patient Advocate

Shawn Martin, MIM
Managing Principal Analyst
Health and Human Services
California Legislative Analyst's Office

Maureen McKennan
Deputy Director
California Department of Managed Health Care

Barbara Mendenhall, MA
Research Program Specialist
California Office of the Patient Advocate

Ed Mendoza, MPH
Manager, Health Information Division
California Office of Statewide Health Planning and Development

Angela Minniefield, MPA[1]
Deputy Director, Healthcare Workforce Division
California Office of Statewide Health Planning and Development

Melanie Moreno, MSW, MPH
Staff Director, Health Committee
California State Senate

Edwin Park, JD
Vice President for Health Policy
Center on Budget and Policy Priorities

Sandra Perez, MHA, MPA
Director
California Office of the Patient Advocate

Senita Robinson
Chief, Research Policy and Planning Section
Healthcare Workforce Division
California Office of Statewide Health Planning and Development

---

[1] Subsequent to her interview for this project, Ms. Minniefield left OSHPD for a position at Charles Drew University.

Janice Rocco, JD
Deputy Commissioner, Health Policy and Reform
California Department of Insurance

Brian Sala, PhD
Acting Director
California Research Bureau

Linette Scott, MD, MPH[2]
Chief Medical Information Officer
California Department of Public Health

Ron Spingarn
Deputy Director, Health Care Information Division
California Office of Statewide Health Planning
and Development

Jonathan Teague
Manager, Health Information Division
California Office of Statewide Health Planning
and Development

Katie Trueworthy
Principal Consultant, Health Committee
California State Senate

James Watkins
Chief, Research and Analytic Studies
California Department of Health Care Services

Meredith Wurden, MPH, MPP
Fiscal and Policy Analyst, Office of Research
California State Senate

---

[2] Subsequent to her interview for this project, Dr. Scott transferred to the Department of Health Care Services.

Appendix B: Sources of Machine-Readable Data on Health Care in California

This chart presents information about each of the machine-readable data sets publically available from state agencies and other sources regarding health care in California.

DATASET	DESCRIPTION	VARIABLES	TIMEFRAME	SCOPE	POPULATION	SAMPLE SIZE	ACCESS	COST
California (OSHPD) Inpatient Discharge Data	Inpatient discharge records for patients in licensed general acute care hospitals in California	Demographic characteristics; diagnosis, treatments, disposition (e.g., discharged to home, discharged to rehab facility), total charges, expected source of payment	1999–2011	State/ County hospitals	Hospital discharges	Approx. 3.6 to 3.7 million per year	Datasets on CDs Downloadable pivot tables* Online query tool*	Public version free from source
California (OSHPD) Hospital Annual Utilization Data	Utilization data from annual reports that licensed general acute care hospitals are required to submit to OSHPD	Occupancy rates, discharges, patient days, types of services provided, expected sources of payment, type of ownership	2001–2011	State/ County hospitals	Hospitals	483 hospitals	Downloadable Excel datasets and pivot tables	Free from source
California (OSHPD) Hospital Financial Data	Annual financial data from reports that licensed general acute care hospitals are required to submit to OSHPD	Type of ownership, number of beds, balance sheets, income statements, revenue by payer, expenses	2002–2011	State/ County hospitals	California hospitals	300 to 400 hospitals per year	Downloadable Excel datasets and pivot tables Online query tool*	Free from source
Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QIs) for California	AHRQ QIs from patient data routinely reported to OSHPD	**Adults:** hospitalization rates for ambulatory care sensitive conditions (e.g., diabetes, heart disease) **Children:** hospitalization rates for ambulatory care sensitive conditions (e.g., asthma, low birthweight) **Inpatient care:** inpatient mortality for select conditions and procedures, rates of select medical and surgical procedures, avoidable complications, adverse events	2005–2011	County hospitals	California hospital and patient discharge data	Varies by year and category of indicators	Downloadable Excel tables	Free from source
California (OSHPD) Emergency Department and Ambulatory Surgery Datasets	Encounter data for visits to licensed emergency departments and ambulatory surgery centers	Demographic characteristics, diagnosis, treatment, disposition (e.g., admitted to hospital, discharged to home), and expected source of payment	2005–2011	State/ County hospitals	Outpatient encounters	100,000 to 450,000 every six months	Datasets on CDs Downloadable Excel tables*	Public version free from source

*Select variables.

DATASET	DESCRIPTION	VARIABLES	TIMEFRAME	SCOPE	POPULATION	SAMPLE SIZE	ACCESS	COST
California (OSHPD) Primary Care and Specialty Clinics Data Products	Data regarding two types of clinics licensed by the state of California: primary care clinics and specialty clinics	**Primary care clinics:** community services, clinic staffing, and patient and staff language data, financial information, and information on encounters by service, principal diagnosis, and procedure codes (CPT codes) **Specialty clinics:** number of surgical operating rooms, number of surgeries performed, number of patients, number of encounters by service type, major capital expenditures	2001–2011	State clinics	Licensed clinics	**Primary care clinics:** 553 **Specialty care clinics:** 1,032	Downloadable Excel datasets and pivot tables	Free from source
California (OSHPD) Long-Term Care Facility Annual Utilization Data	Annual utilization data derived from reports that long term care facilities are required to submit to OSHPD	Ownership, bed classification, patient demographics, patient days, and census by payer categories	2001–2011	State long term care facilities	California long term care facilities	1,100 to 1,220 long term care facilities per year	Datasets on CDs Downloadable Excel datasets and pivot tables	Free from source
California (OSHPD) Long-Term Care Facility Annual Financial Data	Annual financial data derived from financial reports that long term care facilities are required to submit to OSHPD	Type of ownership, number of beds, balance sheets, income statements, revenues by payer, expenses	1997–2011	State long term care facilities	California long term care facilities	1,100 to 1,220 long term care facilities per year	Datasets Pivot tables	Free from source
California (OSHPD) Home Health Agencies and Hospice Facility Annual Utilization Data	Annual utilization data derived from annual utilization reports that home health agencies and hospices are required to submit to OSHPD	Visits, services provided, gross revenue	2001–2011	State home health agencies and hospices	California home health and hospice agencies and utilization	1,600 home health agencies per year	Downloadable Excel datasets	Free from source
Medi-Cal Drug Utilization Data	Monthly and quarterly utilization data for outpatient drugs reimbursed by Medi-Cal on a fee-for-service basis	National Drug Code numbers, strength, dosage, quantity and days supply dispensed, number of claims paid, total amount paid, and amount paid per prescription	1996–2012	State	Outpatient drug utilization	Varies with the volume of prescription claims for Medi-Cal recipients	Downloadable datasets	Free from source

DATASET	DESCRIPTION	VARIABLES	TIMEFRAME	SCOPE	POPULATION	SAMPLE SIZE	ACCESS	COST
California Department of Health Care Services (DHCS) Data for Research and Public Health	DHCS provides protected data (i.e., data containing confidential information) for research and public health purposes. Each request must be approved by DHCS's Data and Research Committee (DRC)	Enrollment, claims, expenditures, demographic characteristics	Varies	State	Medi-Cal enrollees	Varies	Custom datasets* Downloadable pivot tables†	Pivot tables free from source; fee charged for custom datasets
Department of Public Health Vital Statistics Data Files	Vital Statistics Data Files compiled from the information reported on birth, death, and fetal death certificates	Detailed demographic information related to the infant, mother, and father (for births and fetal deaths) or decendent (for deaths), medical data related to the birth or death	Birth files: 1960–2011 Death files: 2002–2005 (Older records available upon request)	State/ County	California births and deaths	**Birth statistical data:** 500,000 per year **Death statistical data:** 116,000 per year	Datasets on CDs* Downloadable Excel tables Online query tool†	Varies across datasets, from $50 to $250 per year of data; online query tools and tables available free
California Cancer Registry	The California Cancer Registry (CCR), California's statewide population-based cancer surveillance system	Incidence and mortality by cancer site, cancer stage at diagnosis, demographic characteristics	1985–2009	State/ Region/ County	All cancers diagnosed in California (except basal and squamous cell carcinoma of the skin and carcinoma in situ of the cervix)	2.5 million cases of cancer to date, with over 140,000 new cases added annually	Datasets provided by agency* Online query tool	Free from source
California Health Interview Survey (CHIS)	A random-dial telephone survey that collects information about more than 50,000 adults, adolescents, and children per two-year cycle. Conducted by the UCLA Center for Health Policy Research	Demographic characteristics, socioeconomic characteristics, health status, health conditions, mental health, health behaviors, access to care, use of health care services, health insurance, participation in public programs, neighborhood, and housing	2001, 2003, 2005, 2007, 2009	State/ Region/ Large county	Households	50,000 persons	Datasets on CD and downloadable from website Online query tool	Download free from source or purchase CD-ROM at $12 per year (adult, teen, or child data)

*Approval required.
†Select variables.

DATASET	DESCRIPTION	VARIABLES	TIMEFRAME	SCOPE	POPULATION	SAMPLE SIZE	ACCESS	COST
California Women's Health Survey	Annual telephone survey that collects information from a sample of 4,000 randomly selected women age 18 and over. Conducted by the Survey Research Group	Past and present involvement in health care systems, food security status, participation in government nutrition programs, prenatal care, vitamin consumption, alcohol consumption, breastfeeding, sexually transmitted diseases, intimate partner violence, utilization of preventive services, demographic characteristics	1997–2011	State	Adult women	4,000	Datasets provided by the Survey Research Group	Free from source
California Environmental Health Tracking Program	Science-based information on the trends and distributions of diseases and environmental threats	Data on hospitalizations and ED visits for asthma and heart attacks, rates of select cancers, rates of select birth defects, infant health indicators	Varies across data sources	State/ County	Varies across data sources	Various sizes and data sources	Downloadable Excel tables Online query tool	Free from source
County and Statewide Archive of Tobacco Statistics	A random-dial telephone survey that includes a screener survey, an adult extended survey, and a youth survey. The adult survey has a sample of 4,000 people. Conducted by UC San Diego and Westat	Smoking behaviors, tobacco cessation, attitudes regarding policies to discourage smoking and reduce exposure to secondhand smoke, attitudes toward the tobacco industry, general attitudes and health beliefs, health consequences of smoking, economic consequences of smoking, exposure to smoking in the media, demographic characteristics	Every three years, 1996–2008	State/ County	Adults and youths	Varies	Online query tool	Free from source
EpiCenter California Injury Data Online	Custom table builder covering all injuries occurring in California to state residents	Deaths, nonfatal hospitalizations, nonfatal ED visits by cause of injury, demographic characteristics	2000–2010	State/ County	Injuries	Various sizes and data sources	Online query tool/ table builder	Free from source
Health Plan Financial Summary Report	Licensing and financial data that health plans are required to submit to the Department of Managed Health Care	Total enrollees, enrollees by health plan type, assets, liabilities, revenue, expenses, administrative cost ratio, medical loss ratio	2002–2012	Health plans	Health plans	Varies by health plan	Online query tool/ table builder	Free from source

Appendix C: Sources of Health Care Data Used by Interviewees

State Agencies

California Department of Finance

California Department of Health Care Services

California Department of Insurance

California Department of Managed Health Care

California Department of Public Health

California Department of Social Services

California Emergency Medical Services Authority

California Employment Development Department

California Managed Risk Medical Insurance Board

California Office of the Patient Advocate

California Office of Statewide Health Planning and Development

Licensing boards
(e.g., Medical Board, Board of Registered Nursing)

Federal Agencies

Centers for Medicare & Medicaid Services

United States Census Bureau

National Associations

National Association of Insurance Commissioners

National Conference of State Legislatures

Philanthropic Organizations

California HealthCare Foundation

Kaiser Family Foundation

Other

California community colleges

California State University

University of California

California
HealthCare
Foundation

1438 Webster Street, Suite 400
Oakland, CA 94612
tel: 510.238.1040
fax: 510.238.1388
www.chcf.org