



SOCIETY OF ACTUARIES

Article from:

Health Section News

August 2005 – Issue 50



Health Section News

"For Professional Recognition of the Health Actuary"

Accessing and Using Public Data: A Primer for the Health Actuary

by Denise Love

"The twenty-first century will be about velocity: the speed of business and the speed of change. ... An infrastructure designed around information flow will be the "killer application" for the twenty-first century."

— Bill Gates, *Business @ the Speed of Thought*

Transformation of the health care industry is occurring and is long overdue. Despite escalating health care costs, health care quality and access have not improved and may be worsening. Pay for performance, patient safety awareness, consumer-driven health care and advancements in medical informatics and information technology are converging to impose change on a health care industry historically resistant to change.

If the 1990s can be characterized as the years of managed care and quality improvement in health care, then the first part of the 21st century might be known as the years of medical informatics and information technology. The good news for data lovers (actuaries, statisticians) is that health data no longer is an afterthought but it is an essential component of health care delivery, payment and decisions. But the next challenge will be information management. Most of us suffer, not from a lack of information, but from information overload. Search engines are more efficient than ever and within seconds deliver thousands of links to Web sites and documents. For example, a Google search of "public health data" delivers in seconds over 193,000 links to agencies, reports and data sources.

The health actuary can benefit from the diverse array of public data sets generated by federal and state agencies. Knowing what types of data are available and where to look reduces time and effort in accessing the right data for the



right task. Knowing where and how to narrow your data search for the right data source can reduce the search time and effort.

About This Article

This article was written by the National Association of Health Data Organizations (NAHDO) for the Society of Actuaries. The paper is a primer for actuaries with limited experience in accessing and using public data sets. First, a very basic inventory of the major federal and state data sets is provided. Next, a few examples of online and analytic tools and innovative Web portals are described. These tools and portals (which organize and point to content created by others) offer a wide range of content appealing to the novice as well as the most sophisticated researcher. Private or proprietary data sets are not included in this article, as many are not available for general public use.

(continued on page 4)

Actuaries

The Best-Kept Secret in Business™

About The National Association of Health Data Organizations (NAHDO)

NAHDO is a national nonprofit membership and educational association, established in 1986. The organization provides technical assistance to and advocacy for public and private health data organizations that collect and disseminate hospital discharge and other health care data. NAHDO regularly convenes leaders in health care information to share best practices and transfer methods and technologies across states and provides formal testimony and consultation to federal and state policy makers around health care data issues. Also, NAHDO is actively involved in national standards development and federal grants and projects to improve the quality, quantity and use of health care data for health care cost, quality and access purposes. Together, NAHDO's senior staff has over 25 years experience in the technical and political aspects of implementing statewide reporting initiatives, including the dissemination of market and policy indicators related to health care cost, quality and access. NAHDO is a subcontractor to Medstat in the Agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project (HCUP) and through a NAHDO-CDC cooperative agreement provides technical assistance and statistical guidance to states disseminating public health data on the Internet. Visit www.nahdo.org.

About Public Data: 101

Because most publicly available data are generally available (for low or reasonable cost) to appropriate users and generally are accompanied by detailed technical documentation, they are an attractive data source for research and special studies. However, the user must be aware of a data set's limitations. Because many public data sets are collected to administer a program or to pay a claim, they may not be designed to perfectly meet the needs of other types of secondary uses, such as research.

Tips on Accessing Public Data

The myriad of online query tools listed later in this article are reducing the barriers to access to federal and state data sets. However, structured queries and aggregate reports are not likely to meet the needs of the serious researcher or actuary who may want to manipulate detailed data. These users will want to access the micro data files directly from the agency, and to do so, need to understand the legal implications of gaining access to the data. While each state has statutes governing the release and use of public health data, the privacy regulations promulgated under the Health Insurance Portability and

Accountability Act of 1996 (HIPAA) have resulted in additional considerations when requesting a data set. Public health agencies continue to experience inconsistent and sometimes conflicting interpretations and application of the HIPAA privacy regulations when dealing with the collection, maintenance, use and disclosure of health information. For bona fide research, HIPAA requires approval by an institutional review board. For non-research applications, HIPAA provides for a "limited data set" with direct identifiers such as name, address, or fields which individually identify a patient removed or masked. A limited data set must also be accompanied by a data use agreement. For detailed information about the HIPAA privacy rule, see the following links:

<http://www.mc.vanderbilt.edu/root/vumc.php?site=hipaaprivacy&doc=1548> and

<http://www.hhs.gov/ocr/generalinfo.html>

Tips for Requesting Public Data

- Structure your data request to avoid delays or getting turned down.
 - Define your study period. How many years of data do you need versus what is available? For multi-year studies, be aware of changes in hospital ownership over the time period. Codes and definitions may also change between years. Data elements may be added or deleted across years, so data documentation is critical. Be aware of calendar year or fiscal year time frames.
 - Consider the universe. Do you need all hospitals in an area? All geographic areas? Are data available for the scope of your study (e.g., specialty hospitals may be excluded from some statewide hospital discharge data sets)?
 - Be aware of legal or other limitations/restrictions to data release and disclosure. Some states restrict the public disclosure of hospital identity, as does the HCUP national inpatient sample. Some agencies limit public disclosure to aggregated results and restrict secondary release of the data.
 - Most hospital discharge data sets release charges, not cost or payment.

Where to Find Public Health Data

Federal Government Data Resources

The Department of Health and Human Services (DHHS) maintains a broad array of data collection systems designed to monitor disease outbreaks, disease treatment outcomes, injuries, food safety and other public health problems. Individual federal agencies are also providing Web tools to increase access to their own statistics and data sources.



Denise Love, B.S.N., M.B.A., is executive director of the National Association of Health Data Organizations. She can be reached at nahdoinfo@nahdo.org.

- Centers for Disease Control and Prevention (CDC) — www.cdc.gov
 - Numerous national-level surveillance data reported by states, ranging from cancer to pregnancy risk assessment data, are maintained by the CDC.
- National Center for Health Statistics (NCHS) of the CDC (www.cdc.gov/nchs)
 - NCHS maintains a host of household and provider-level surveys
 - National Health Interview Survey (NHIS)
 - National Health and Nutrition Examination Survey (NHANES)
 - State and Local Area Integrated Telephone Survey (SLAITS)
 - Survey of Income and Program Participation (SIPP)
 - National Employer Surveys
 - National Immunization Survey (NIS)
 - National Maternal and Infant Health Survey (NMIHS)

NCHS and other DHHS agencies also conduct provider-level surveys that collect data from hospitals, physicians and clinics. Some of these surveys collect information directly from the individuals who use these services, but all of them also collect data from facility records.

- National Ambulatory Medical Care Survey (NAMCS)
- National Hospital Discharge Survey (NHDS)
- National Home and Hospice Care Survey
- Agency for Healthcare Research and Quality (AHRQ) — www.ahrq.gov
 - National Medical Expenditure Panel Survey (MEPS)—An annual household survey conducted since 1996 using the NHIS sample frame.
 - Consumer Assessment of Health Plans Survey (CAHPS)
 - Healthcare Cost and Utilization Project (HCUP)
- Substance Abuse and Mental Health Services Administration (SAMHSA) — oas.samhsa.gov
 - Drug Abuse Warning Network (DAWN) from hospital emergency department records
- Centers for Medicare and Medicaid Services (CMS) formerly HCFA — www.cms.hhs.gov/researchers
 - Medicare program data are widely used to study health and health care outcomes of populations eligible for Medicare coverage. The Medicare Enrollment Database (EDB) contains information on all Medicare beneficiaries. It is an important database because it can link to

other Medicare files. Medicare Current Beneficiary Survey Series (CMS) and the Medicare Provider and Review (MEDPAR) files.

- Health Resources and Services Administration (HRSA) — www.hrsa.gov/data.htm
 - HRSA provides a wide range of data and statistics on maternal-child health, workforce, primary care, rural health and health insurance coverage.

Other federal data:

The Census Bureau, part of the U.S. Department of Commerce, is one of the primary sources of insurance data; it conducts two main surveys responsible for deriving health insurance data—the Current Population Survey and Survey of Income and Program Participation.

Human Services data include the Temporary Assistance for Needy Families (TANF). States provide data on a quarterly basis to the federal government including data on employment, earnings and income from other sources.

Federal Portals

- Quick Access to Federal Government Data (<http://www.fedstats.gov/>)
This site is a gateway to statistics from over 100 U.S. federal agencies and provides direct access to federal agencies, online data resources, mapping statistics and almost any federal statistical resource.
- HHS Data Council Gateway to Data and Statistics (www.hhs-stat.net/)
This Web-based tool brings together key health and human services data and statistics. It is designed to complement other government resources such as FirstGov and FedStats. The Gateway covers federal, state and local government sponsored information.

State Health Data

States are responsible for maintaining numerous health-related data collection systems including vital statistics (birth and death records); hospital discharge abstracts which provide detailed information on hospital patients and the diagnoses and treatments they receive; registries such as the cancer registry system; and programs such as Medicaid and State Children's Health Insurance Program (SCHIP). Much of the data states collect are shared with DHHS for department use in monitoring the health of the nation and administering and evaluating federal programs.

Because states' regulatory powers and service provision activities are broad, the federal government relies on states to collect health data used to study health and health services at the state and federal levels.

Much of the data resources are located with state health departments. The most efficient way to access one or more health department home pages is through the CDC Web site at www.cdc.gov/mmwr/international/relres.html which lists each state health department.

Important federal-state cooperative data initiatives reflect the critical data partnerships between the federal and state governments, where the state implements data collection and management using federal guidelines and standards, and then reports local data to the federal agency. Examples of these cooperatives and partnerships include the following:

- **Vital Statistics Cooperative Program:** The National Vital Statistics System is the oldest and most successful example of inter-governmental data sharing in public health and the shared relationships, standards and procedures form the mechanism by which NCHS collects and disseminates the nation's official vital statistics. These data are provided through contracts between NCHS and vital registration systems operated in the various jurisdictions legally responsible for the registration of vital events—births, deaths, marriages, divorces and fetal deaths. Visit <http://www.cdc.gov/nchs/nvss.htm>.
- **Statewide Hospital Discharge Data Programs:** Over 45 states maintain statewide, discharge data systems that include all payers on all patients admitted to acute care hospitals, including the uninsured. These systems are maintained by state agencies or private data organizations, such as a hospital association. All of these data programs collect inpatient data in a uniform billing 92-based discharge data abstract which may be modified by states to meet local needs. Data access policies vary by state, depending on the legal and organizational policies governing data collection and release. Many of these states also participate in the HCUP project, which provides state-level data files in a uniform or HCUP format. Generally, the state agency provides research-level or more detailed data guided by data use agreements and policies. Many states are disseminating hospital statistics in query format on the Internet. Contact NAHDO at nahdoinfo@nahdo.org for contact information.
- **Healthcare Cost and Utilization Project (HCUP):** HCUP is a family of healthcare databases and related software tools and products developed through a federal-state industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP databases bring together the data collection efforts of state data organizations, hospital associations, private data organizations, and the federal government to create a national information resource of patient-level health care data. HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs and outcomes of treatments at the national, state and local market levels. More information, databases and tools are available at <http://www.hcup-us.ahrq.gov/overview.jsp>
- **Behavioral Risk Factor Surveillance System (BRFSS):** The BRFSS is a state-level survey developed by DHHS in collaboration with the states to monitor state-level prevalence of behavioral risks among adults. The survey contains a core survey that is common across all states so that comparisons can be made, but flexibility to permit states to add their own questions.
- **Youth Behavioral Risk Factor Survey (YBRFS):** A state-level survey modeled after the BRFSS and targeting adolescents.
- **Surveillance, Epidemiology and End Results (SEER) program for cancer:** The National Cancer Institute administers the SEER program to provide data on cancer incidence and survival. Data are collected from cancer registries in 14 geographical areas covering approximately 26 percent of the U.S. population
- **Medicaid and State Child Health Insurance Programs (SCHIP):** States report encounter data to the CMS Medicaid Statistical Information System (MSIS). MSIS data are used to create an analytic data file, which prior to 1999 was called "SMRF" but now is named "Medicaid Analytic Extract" (MAX). MAX files include claims and encounter records in a revised format. MAX files include encounter data from managed care organizations, but CMS staff does not consider these data to be useful for research purposes, as discussed below. For each state for each year, there are five MAX files, an eligibility file plus four utilization files (the same types that states use

when submitting their data to CMS). Researchers outside of the federal government can purchase these files for approved research activities through a data use agreement with CMS.

- **State Health Interview Surveys:** The national surveys do not support state or local estimates, so many states conduct their own state-specific surveys, and about 25 states have received federal funding from HRSA to conduct state planning grants to study potential ways to expand health insurance. The goal of the program is to support states as they analyze their uninsured populations and healthcare marketplaces in order to develop solutions to ensure health coverage for all state residents. More information can be found at <http://www.hrsa.gov/osp/stateplanning/granteelist.htm>.

The California Health Interview Survey (www.chis.ucla.edu/) is an example of a state health interview survey.

Public Domain Analytic Tools

National measures of quality increasingly used for proprietary, purchasing, public reporting and quality improvement initiatives are the AHRQ Quality Indicators. AHRQ's Quality Indicators are standardized indicators of quality generated from widely available hospital discharge data sets.

The quality indicators were empirically evaluated and refined by Stanford University's Evidence-based Practice Center. Under contract with AHRQ, Stanford assesses the face validity, precision, bias, construct validity and application factors for each quality indicator. This study resulted in the development of three software modules. The software can be downloaded without charge at www.qualityindicators.ahrq.gov/ in SAS or SPSS format (and soon an online calculation tool will be available).

The advantages of the indicators are their public access, complete documentation, standardized definitions and a reference database consisting of 35 state inpatient data sets. The indicators can be used with any hospital administrative data set, including MedPar, state discharge data, payer data and a hospital's internal data. Known limitations of administrative data apply to these indicators as they do to any study involving billing or claims data. Each of the following software modules generates numerators, denominators, observed rates, risk-adjusted rates and smoothed rates for individual indicators.

- **Prevention Quality Indicators (PQIs)** are a set of 16 measures that can be used with hospital inpatient discharge data to identify quality of care for "ambulatory care-sensitive conditions." These are

conditions for which good outpatient care can potentially prevent the need for hospitalization or which early intervention can prevent complications or more severe disease.

- **Inpatient Quality Indicators (IQIs)** are a set of 31 measures that provide a perspective on hospital quality of care using hospital administrative data. These indicators reflect quality of care inside hospitals and include inpatient mortality for certain procedures and medical conditions; utilization of procedures for which there are questions of overuse, underuse, and misuse and volume of procedures for which there is some evidence that a higher volume of procedures is associated with lower mortality.
- **Patient Safety Indicators (PSIs)** are a set of 29 indicators to help health system leaders identify potential adverse events occurring during hospitalization. Twenty-three of the PSIs are provider-level measures and six are area-level measures. The PSIs are a set of indicators providing information on potential in-hospital complications and adverse events following surgeries, procedures and childbirth. The indicators can be used to help hospitals identify potential adverse events that might warrant further study.

Web Query Systems to Disseminate Public Data

State and federal agencies are developing interactive Web query systems to disseminate health statistics on the Web. These sites provide a quick and easy way for researchers to assess the significance of a problem and explore the data prior to purchasing the entire data set for detailed studies. It provides consumers and advocacy groups with aggregate information about a particular condition or procedure. And they can be used to gather national or regional benchmarks for use with local or proprietary data sets. A more complete listing of national, state and local Web query systems can be found at NAHDO's Web site, the Health Information Dissemination Systems Clearinghouse (HIDSC) at <http://www.nahdo.org/hidsc2/hidsc.aspx?id=Users%20web%20applications>.

Table 1 Web Query Tools, Selected Examples

Entity	Name	Description/Criteria	URL
Utah Department of Health	Indicator-Based Information System for Public Health (IBIS-PH)	<p>This system contains standard reports, publications and multiple query modules, which access data on population estimates, births, mortality, hospital use, emergency department use, health surveys, cancer registry and injuries.</p> <p>Emergency department module.</p> <p>Metrics: counts, crude rates, age-adjusted rates, total charges, average charge and median charges.</p> <p>Filters and dimensions: year, diagnosis, procedure, gender, primary payer (including Medicaid, SCHIP), discharge status and geographic area.</p>	<p>http://ibis.health.utah.gov/view?xslt=home.xslt&xml=home/home.xml</p>
Wisconsin	Wisconsin Inquiry Tool for Healthcare Information (WITHIN), Ambulatory surgeries query module	<p>WITHIN, which is based on Utah's IBIS-PH system, allows queries of hospitalizations and ambulatory surgeries (from both hospital-based and freestanding facilities).</p> <p>Ambulatory surgeries query module.</p> <p>Metrics: counts, total charges, average charge and median charges.</p> <p>Filters: type of surgery (170+ options) gender, age group, county of residence, year and primary payer (including medical assistance).</p> <p>Dimensions: year, gender, age group, county of residence and primary payer.</p> <p>Years available: 2001 and 2002.</p>	<p>http://dhfs.wisconsin.gov/within/qspages/qcamb01.htm</p>
South Carolina	Analysis of Emergency Room Discharges by Selected Characteristics	<p>Metrics: total and average charges.</p> <p>Filters: diagnosis category, specific diagnosis, age group, race, gender, primary payer (including Medicaid), county of residence, health service area and health district.</p> <p>Dimensions: county of residence, health service area, DHEC health district and primary payer.</p> <p>Years available: 2002 and 2003.</p>	<p>http://www.ors2.state.sc.us/er.asp</p>

Entity	Name	Description/Criteria	URL
West Virginia	Health IQ 2003	<p>Metrics: number of hospital discharges, charges, inpatient days, average charge and length of stay.</p> <p>Filters and dimensions: gender, age group, county of residence, payer, type of service, discharge status, DRG, APS, MDC, principal and secondary diagnosis and principal and secondary procedure.</p> <p>Years available: 2000-2002.</p>	http://www.hcawv.org/DataAndPublic/IQ/UB03.asp
AHRQ	Healthcare Cost and Utilization Project (HCUP-NET)	<p>HCUP-net generates statistics using data from HCUP's Nationwide Inpatient Sample (NIS), Kids' Inpatient Database (KID) and State Inpatient Databases (SID).</p> <p>Metrics: number of discharges, mean and median length of stay, mean and median charges, percent died in the hospital, discharge status, percent admitted from emergency department, percent admitted from another hospital and percent admitted from long term care facility.</p> <p>Soon will include the AHRQ Quality Indicators statistics.</p>	http://hcup.ahrq.gov/HCUPnet.asp
AHRQ	Medical Expenditure Panel Survey (MEPS)	<p>MEPS has two components: household and insurance. Household component:</p> <p>Metrics, filters and dimensions: hospital emergency room visits, prescribed medicines, perceived physical and mental health status and insurance status.</p> <p>Years available: 1996-2002.</p>	http://www.meps.ahrq.gov/mepsnet/mepsnetintro.htm
NAHDO	Emergency Department Internet Query System (EDIQS)	<p>This query system provides national emergency department statistics and benchmarks derived from the NCHS National Hospital Ambulatory Medical Care Survey (NHAMCS). Users can query general and injury-related ED utilization statistics by patient and hospital characteristics.</p>	<p>Available at the NAHDO site: http://155.98.221.34/ediq/index1.htm</p>

Table 2 Examples of Model Local Web Resources

Local Portal	Characteristics	URL
Massachusetts Health Data Consortium	Catalogues links to health data sites by: <ul style="list-style-type: none"> • Costs/expenditures • Disease/conditions • Drugs • Facilities • Geographic • Insurance • Medical Care/Treatment • Health Care Workforce 	http://www.mahealthdata.org/
Health Foundation of Greater Cincinnati	Maintains a Health Data Resource directory for the Tri-State area of Indiana, Kentucky and Ohio. Created the Online Analysis and Statistical Information System (OASIS) in partnership with the University of Cincinnati. OASIS permits user-defined analysis of data sets in its data warehouse for guided analysis or execution of sophisticated statistical functions. Mapping software permits the generation of maps. SAS logs are generated and downloadable, as are data sets. Detailed documentation of codes and data fields are available for data sets in the warehouse.	http://www.healthfoundation.org/data OASIS: http://www.oasis.uc.edu/OASIS_CODE/Templates/Login.cfm
Family Health Outcomes Project, University of California San Francisco	This site includes excellent information about data, and online access to public health data through FHOP-maintained interactive sites. <ul style="list-style-type: none"> • EpiBC 2005: birth certificate data • Analysis and presentation system • Hospital discharge data analysis and presentation system • EpiMap2 california county map boundary files Downloadable EpiInfor (ver 3.2.2) with full users manual	http://www.ucsf.edu/fhop/htm/pub_health_data/index.htm

Local Portal	Characteristics	URL
<p>Washington State Department of Health</p>	<p>Health Data Section: links you to pages within and outside the Department of Health Web site that contain links to data tables or data for online query and publications. VistaPHw is used across the Washington State public health system as a standardized tool for community health assessment.</p> <p>Statistical guidelines for commonly encountered issues in public health practice. Assume a basic knowledge of epidemiology and biostatistics.</p> <ul style="list-style-type: none"> • Confidence intervals for public health assessment • Population denominators • Racial and ethnic groups in data analyses • Rates for public health assessment • rural-urban classification systems for public health assessment • Small Numbers • Address matching and geocoding data • Human subjects review <p>Many health publications.</p>	<p>http://www.doh.wa.gov/Data/data.htm</p>
<p>University of Michigan's Statistical Resource on the Web for Health</p>	<p>Regularly updated, provides data and statistical resources for topics ranging from A to V (Abortion to Vital Statistics). Statistical Universe indexes and abstracts federal government statistics since 1974; business, association, and state government data since 1980, and international agencies since 1983. About 15 percent of the abstracts link to full text.</p>	<p>http://www.lib.umich.edu/govdocs/sthealth.html</p>
<p>National Association of Health Data Organizations (NAHDO)</p>	<p>The NAHDO-CDC Cooperative Agreement supports a Web site, Health Information Dissemination Systems Clearinghouse (HIDSC) with links to interactive public health Web sites, plus:</p> <ul style="list-style-type: none"> • Statistical guidelines • Soon HIPAA white papers series • Technical papers series 	<p>http://www.nahdo.org/hidsc2/hidschome.aspx</p>

The actuarial community is an important constituency or user group for federal and state data sets. Since these data systems rely on public funding, actuaries can help. Often legislators will want to know who uses the data and its benefits. If you use a public data set for a study or in your

daily work, provide feedback to the agency about the data, what might be improved, and results or findings from your study. This information is helpful to agency staff, especially as they prepare for their budget or sunset reviews. 📧