

RECORD, Volume 29, No. 2*

Spring Meeting, Vancouver, B.C.
June 23–25, 2003

Session 123TS

Public Databases and Other Resources For Health Actuaries

Track: Health

Moderator: RICHARD A. KIPP

Instructors: RICHARD A. KIPP
DENISE LOVE†

Summary: This teaching session provides a summary of publicly and privately available data and information sources for health actuaries. Several of the more useful resources are reviewed in greater detail, including detailed instruction on the SOA's data search engine.

MR. RICHARD KIPP: It's always of importance for actuaries to have good, sound, solid data sets to work with, build models and do analyses. Data obtained from public and private sources can be used to complement the data and the analyses that we would be able to perform using data available from internal sources. Actuaries are in a constant search for data. Recently I was re-appointed to the Milliman USA Health Steering group, which is Milliman's national health discipline organization. That group focuses on obtaining data and doing research corporately. Among other things, I am now primarily responsible for maintaining our data sets and finding new data. It is a very extensive process for us to obtain, renew and refresh the data that we get from our various sources.

Denise Love is also here to present. She is the executive director of the National Association of Health Data Organizations (NAHDO).

It is amazing, really, when you think about it. Since 1998, when I last gave this talk, things have changed dramatically in terms of the data that is accessible through the Internet—especially data that is offered by the variety of companies that are interested in not just data capturing and selling, but also in decision support, etc. A great deal of the data that is offered on the market today on the private side comes to those organizations by virtue of the fact that they supply decision support services to their clients. We will look at some of those data sets.

* Copyright © 2003, Society of Actuaries

†Ms. Love, not a member of the sponsoring organizations, is executive director at National Association of Health Data Organizations in Salt Lake City, Utah.

On the private side, you will find that some of the available data is rather pricey. It depends on what you want and how much you want. I will not get into the pricing issues, but we will be pointing you to Web sites that you can go to for various types of data that are accessible, both on the public and the private side.

Deciding what external public and private data that would be of interest to actuaries was something that I struggled with. At some basic level, to the extent that you work for an insurance company or a payer of one sort or another, you probably have access to a wealth of information—fairly detailed data. I know that's not always true. Certainly years ago, when I worked directly for Plans, it wasn't always the case that the actuaries were given access to that detailed data to do their analyses. That has changed with time.

Actuaries, in general, are given access to their own company's data, a little more readily than they were in the past. The notion of the data "people" and the data "shop" keeping the data away from the end user has sort of disappeared. Lack of direct access to data was a substantial barrier to actuaries that were doing detailed analysis back in the day. It still depends on the company that you are dealing with. Some have much more open access and much cleaner and readily available data for actuaries to use than others.

I will discuss general issues of data, reasons for data and types of data, before we actually get into the sources. I have a few sources that I have used from time to time. I will go through some of that private data, as I mentioned.

You might find, as actuaries, common reasons for the need for data—and again, this would not necessarily be external data. This could be your own internal data. If you're going to do a forecast or prepare claim cost estimates or revenue projections for a budget, you're certainly going to work with your own data most often. When it comes to things like provider contracting support, you might want to search outside the organization to find what other data exists that might be used to help create satisfactory contract terms and achieve the best reimbursement levels.

When it comes to rate filings, you may want to go to the state Web sites. Take a look at some rates that have been filed by insurance companies with insurance departments. There are some states that publish that sort of information at one level or another. Forecasting trends will require access to your own data, but you may want to look outside your own system to see what, if any, other data is available that would complement and supplement the data that you have internally.

When it comes to product development and marketing support, marketing people come to actuaries constantly to have new products developed. A lot of that work will depend on access to fairly detailed data that you may have internally. It also may require you to look outside of the organization to find other data. Performance monitoring is one of those benchmarking exercises that you might undertake, for which you will have to have a notion of what performance measures are generally being used by the industry.

You will need to look outside of your organization to find data that allows you to compare yourself to the rest of the industry. More data and services are available along those lines today than there have been in the past. Many ad hoc questions that come up invariably make their way to the actuarial department, and a number of those require outside data. To perform most of these studies you really have to have the skills and the software available to dig into some detailed and large data sets.

The product development area relies a bit more on an outside look. In addition to using your own data to do such things as crafting probability distributions to evaluate things like a consumer-directed health plan product, having access to outside data can be critical to creating a credible basis for analysis. Companies that are going to be most interested in external data are companies that have little or not very credible data, because they have either never sold a particular type of product before or they are new to a particular marketplace. The general need to compare your own company's information, like the claim cost that you're developing in a product for a market, is something that is always present for the actuary.

Organizations are always wondering if there is something actuaries can do to improve the company's results. Is there a new managed-care program that can be put into place like disease management? Is there a new product that can be put into place to improve financial results? The general need to compare is constantly there for actuaries.

Quality is an issue on the rise. There are a number of quality measures and data that are published with regard to quality that you can compare yourself to. Obviously, that requires developing those same statistics for your own plans and being able to make those comparisons so that you can see how certain providers may stack up in various markets. There are many efficiency tools around that allow you to compare lengths of stay or cost for a unit of service, and so forth. To the extent that you have established discounts in various markets, some data sources exist that allow you to test yourself.

You may not find out precisely what your competitor's discounts are from some of these data sources, but you can test your organization's fees to see how you fare in any particular market with regard to some of your contract rates. General market research, rate comparisons and benchmarking of information are done. Quality information is available at the provider level now, as well, depending on the state that you happen to examine.

I am not going to explore every Web site that is available, but I'm going to talk about a few of them like HealthGrades (www.healthgrades.com), just to show you some of what is available. HealthGrades is not always the most detailed, in terms of the scores for the reported measures, but if you go to the site and research the State of California, you can look up coronary artery bypass graft information. You can select a hospital area. You are going to be able to see a contrast of some of

these statistics at a fairly high level. HealthGrades uses a "star system." That is, a number of stars are given to indicate the score for a particular measure. You are not getting detailed values for the particular statistics that are mentioned, but you can see that in the Los Angeles area there are a number of hospitals that are represented, and various quality statistics run across the column headings.

The number of cases for coronary bypass surgery is an important statistic. It is a statistic that LeapFrog Group (www.leapfroggroup.org) has focused on a bit. Hospitals that do more coronary bypass surgeries tend to have better quality results—at least reportedly better quality results—than those that do fewer of them. In this situation, you have quality measures—in-hospital mortality at the time of the surgery, within a month afterward and six months afterward. The Web site gives you an opportunity to click on some LeapFrog Group results, and so forth. I think we will find that there is going to be a lot of focus on quality measurement. Actuaries are going to be required to go into data sets and calculate the values for some of these measures—measures like the HealthGrades measures.

You probably have heard of the National Quality Forum (www.qualityforum.org). Their Web site is a little tricky to navigate, and they do not have what I would call data, per se, but they do have some interesting information about the measures themselves. They list a number of the measures that they are finding to be of importance. The table of contents indicates what measures are highlighted.

For instance, for acute coronary syndromes, a measure is aspirin at arrival for acute myocardial infarction (AMI). It gives a description of what the expectation is. So these are indicators on a given case that would be looked for, captured, accumulated and reported—if The Quality Forum has its way—so individual organizations can be compared with regard to their effectiveness in doing some of these quality measures. I think that it is going to be hugely important for us as actuaries to be aware of these things, to understand the variance around these measures.

There has not been much discussion in most of what I have seen so far on the topic of quality measures about variance in individual measures. There is a lot of data reported or beginning to be reported on the average statistic—the point estimate, if you will, for a particular measure and the value for that measure. But we really don't have a sense of the variances around those measures to test whether or not one hospital's result is truly an indicator of one institution being better or worse than another when compared to another hospital or the state's results. I know that the Hospital Cost Utilization Project (HCUP) analysts have been working on that. There is a little more statistical science going into the development of those measures.

In terms of utilization standards, there is a couple that I could readily find. I am sure that there are many of these types of tools embedded in various software products. Many of those products have a method of benchmarking embodied in them, but they do not necessarily sell the benchmarks as a separate product. You

buy the decision support tool, and then, as a by-product of buying that tool, you have access to some of their benchmarks. That is not as satisfying for an actuary who wants to buy only some benchmarks and have a notion of what's happening with these statistics.

Provider contracting support has various issues. It is probably toughest to know your competition's discount. Network performance, measured one way or another, would be something that might be available. You might get statistics about your competition with regard to the breadth and depth of their network, the types of specialties, etc. You may be able to get some notion of fees by doing background checks with physicians that belong to overlapping networks so you may be able to unravel the mystery of fee levels for procedures that some providers offer. But generally, having access to an entire fee schedule for a competitor is very uncommon.

In terms of capitation, there are some reports. There are surveys that are done from time to time that give you the ability to get a sense for capitation levels. Capitations for a particular specialty product—X-ray, lab, oncology caps, things of that sort—are reported through the Warren Survey (www.demarcowarren.com). There's another publication called *Capitation Rates and Data* that does a survey in which they capture information about capitations. However, they're not that good about giving you the details on the scope of those capitations.

FROM THE FLOOR: What about reasonable and customary schedules?

MR. KIPP: With regard to fees themselves, again, while there is not much good data about competitive fee schedules, you can buy data from Ingenix. They're one of the major data providers that exist today. They have bought up a number of companies over the last couple years, not the least of which is St. Anthony's. And I believe that it was through that St. Anthony affiliation, and purchasing MediCode, that they became the premier data provider with regard to fees.

Ingenix gives access to charge-level data, which is probably a little less interesting than reimbursement levels. They also sell access to a file called the "Allowed file," which does provide physician reimbursement levels. It gives, in three-digit zip-code area-level detail by procedure code, a fee level in percentile ranking. If you take your fee schedule and compare it to a particular marketplace for a particular zip code area, you can get a sense of how your discount compares to the reported discount in a particular marketplace.

That's probably the most detailed fee analysis that you could do on the professional side. There are always questions about the credibility of the data that they're getting. I mean, some of that is coming through surveys. Some of it is coming through the cooperative relationships that they have with some of their clients. So they get access to fairly detailed fee-level data by virtue of having access to claims data that they get from their clients, but a lot of it is based on surveys.

So you have to worry about the accuracy and credibility of the fees for the minor service areas. Again, it is probably one of the more detailed and interesting sources from which to get a sense of how you are stacking up. With regard to risk-sharing targets, if you were trying to compare yourself to other organizations, you would have to have copies of the contracts that they are getting with their other provider organizations to be able to get a good idea of how your terms may be comparing with their terms.

Product development—whether or not you're trying to calculate a new benefit price, benefits and other relativity factors, a carve-out or some special scope of service rate per month per member (PMPM)—if your data doesn't get to that level, purchasing some of these data sets are a way of getting access to detailed information to help you with those projects.

Some of the recent products that have been on actuaries' minds have been point-of-service or PPO products. Also, consumer-driven health plan business certainly has captured the imagination. Whether or not it captures the market share remains to be seen. You may be able to use some of this detailed data to the extent that you do not have access to data in a particular market that you would like to go into. Purchasing some of this private data would give you a way of doing modeling that would help you price a product like one of these consumer-driven health plan products.

Generally, the types of data that seems to be available are: cost, utilization and performance. You can get financial information. InterStudy (www.hmodata.com) actually sells financial statement information for HMOs, in particular. You can get most of the schedules that are in an annual statement from that particular data source. There are also some state Web sites that are publishing financial statements, at which you can get access to financial performance for many organizations.

The actuary would get some idea about dollars per unit of service by provider type, making some assumptions about discount if he were doing some pricing for a new marketplace. These are things that the actuary is going to have to be worried about at a fairly detailed level—mix of procedures, mix of providers—all the elements of a model that you might develop to calculate a price, whether it's the utilization or the population differences.

That brings us to the sources to which you might go to try to find some this data. And as I mentioned, there are public data sets available and more are on the way that can provide interesting insights. There is a number of private and some published data that we'll review very briefly

The federal government has a great deal of information at www.hrsa.gov. They sell detailed claim information. You may have worked with some of it. They actually have a sample for Medicare recipients that has claimant-by-claimant information for

5 percent of the Medicare population. You can track individuals longitudinally so that you can do disease-management studies.

There are some Health Insurance Portability and Accountability Act of 1996 (HIPAA) issues. As a business associate, for any data that we are getting from our clients, we have to worry about HIPAA. The federal government has taken the position that they're going to be a little more restrictive in letting organizations have access to their detailed data—in particular, the 5 percent sample, but also claim data at a more summary level.

The federal government does not have the same sort of data set for all Medicare beneficiaries for hospital outpatient quite as readily available. But in any case, for all of this detailed data, organizations like ours that use it for research have to fill out an application that tells CMS, in a fair amount of detail, what the intended use for that data is going to be. It seems reasonable on the surface, but you have got to do that every time you want to run a program on that data. So if we want to do 20 different studies and cut the data different ways, theoretically, for every one of those studies, we would have to fill out another application and ask for permission to use the data.

This is what the stated policy is at this point. We're not exactly sure how the government is going to administer it. Our organization already has applied, and we are trying to work through some of the red tape that seems to be developing as a barrier to getting access to some of this information.

I often go to www.insure.com to get at all state insurance department information through one site. You can get rate quotes, if you wanted to use it for that. I use it as a "state gateway." In the middle of its Web page, there is a drop-down box that you can use to select a state. It will bring up the Web page for that state and give you the latest news. You get access to different information, depending on the state. Most states will give you access to regulations and so forth.

There is also a search engine that was created for the Society by virtue of a grant that the Society let to the National Association of Health Data Organizations (NAHDO) several years ago (www.nahdo.org).

Some of the data that we constantly use is accessible on the government Web site, such as the adjusted average per capita cost (AAPCC) for a lot of our Medicare + Choice business, and then MEDPAR, BMAD and the 5 percent sample use CMS.gov.

Some of the private data sources are the American Association of Health Plans (www.aahp.org) and Blue Cross and Blue Shield Association (BCBSA) (www.bluecares.com). Ingenix's site (www.ingenix.com) gives you access to information. That site has a lot of different aspects to it and a lot of different data. Ingenix actually has actuarial consulting that it has purchased over the course of the last couple years, Reden & Anders Ltd. Also, Tillinghast's health group has been

acquired by Ingenix, which also has an affiliation to United Health Care. Reden & Anders offers some of the data that you may be interested in.

Thompson Publishing Company is the organization that has purchased Medstat (www.medstat.com) and is a holding company for a number of other data companies that offer data that might be used by the health actuaries.

Medstat offers information about its various client market products, provider, health plan, etc. You can click on products and services and detailed data. They have fairly detailed claims data. In addition to some of these other things, they sell the "Medstat Market Scan Database." It represents 11 million covered lives. They obtain that information in a cooperative agreement with employer groups that they work with and provide decision support to. They have gotten permission from all of these employer groups to package the data together. It is as detailed claim data as you can find on the market these days. They have enrollment files that give you an idea of what type of benefits are in place for the covered individuals, and they give you the effective dates for each individual for the benefits that they have. On the claims database, they give you provider-level detail. They do not identify the providers in any way that you could really make sense of, but they do identify diagnosis codes and common procedural technology (CPT) codes. You can crudely identify individuals, although the information has been de-identified. This information is detailed enough that if you want to complement your own organization's detailed claims data, you could purchase a data set from this company, and you'd be able to work with it to produce claim-probability distributions. You can look at trend longitudinally, changes in behavior with regard to certain diseases, etc.

If you are checking credit, you might want to use Dun & Bradstreet's data. The Warren Survey data that I mentioned before is useful. Reden & Anders is one of the big four or five private data organizations that would be important for you all to know about. They have this actuarial toolbox that does many different things. They have a research database.

I cannot tell you much about the Reden & Anders database, because this is one that I have not purchased. This particular database claims to go from July 1997 through December 2000. The company has complemented it with inpatient databases from public sources. This is data that it has obtained in some cooperative fashion with either its clients or its parent company. And they use it for research purposes. It feeds into some of its actuarial toolbox products. I'm not sure whether they would literally sell you this data at this very detailed level, like Medstat would sell you the data. My guess is that if you had a special study in mind, you would certainly be able to buy access to part of it in some fashion. It may be through one of their software packages.

The other data that Reden & Anders apparently has access to is the Medicare 5 percent sample, the MEDPAR data, the Florida hospital discharge data, some of these other discharge data sets. And the company has several years of that data

available to its researchers. Whether it sells access to that, I'm not exactly sure, but they are one of the big data players.

I will not bring up the Milliman page (www.milliman.com). There is a Web site, and there are tools sold by Milliman that use its data from behind the scenes. Their health-cost guidelines, and so forth, rely on all of the detailed data that the company obtained either by purchasing Medstat, by cooperative agreements with clients, or the hospital-discharge data that the company buys from the various states that sell it.

The Employee Benefit Research Institute (www.ebri.org) has interesting reports on various topics. The BCBSA, like the American Association of Health Plans (AAHP), has reports from time to time that offer very interesting information about trends.

InterStudy data is sold in various ways. There is an HMO financial performance report. There are competitive-edge reports that give you market research information about various markets that you might be interested in, if you're not already selling product in that area. In general, it gives historical data about what has been happening in the HMO industry.

AIS (www.aishealth.com) tends to publish reports on things like disease-management protocols that could be interesting if you are doing modeling of diseases for a disease-management product.

I mentioned client-indemnity data and degree of health care management. To the extent that we at Milliman purchase external data, we have to have some sense of what the underlying data is to be able to use it properly. For some of these data organizations, it is a little tricky to do that, because they sell aggregations of very diverse data sets. Work with their sales team and their product support people to get as good an idea as you can of what constitutes that data. It can be tricky to use, and it could be totally inappropriate.

IMS America (www.imshealth.com) sells data. Milliman USA buys it to work with its health cost index report to do trend analysis. IMS sells survey data on physician-visit volumes, and various survey-based things like that. The pharmacy data that they have is quite detailed. It represents volumes of drugs that are sold to pharmacies, and therefore, it can be used to track the utilization of pharmacy benefits. Milliman has noticed recently that drug trends seem to have dropped off a bit, and this data provided us with insight into this matter.

McKesson HBOC (www.mckessonhbc.com) tends to produce decision-support tools. So again, benchmark information and other data that you would have access to through HBOC would probably be offered through its decision-support software.

The Kaiser Family Foundation (www.kff.org) produces reports and surveys that are used constantly by researchers. It is good to be aware of what's going on in those reports, because you may have to battle a broker or an employer group that is also

using it. They may be looking at these reports and reading about trends and what is happening in the marketplace. When you are doing your pricing and developing your new products, you may have to defend against some of the information that's published there. Not that it is wrong, but it is a publicly available tool that a lot of people are looking at in the employee-benefit world. It offers detailed charts and tables. It tracks trends. It tracks premium trends. The Foundation has been doing its survey for a long while. It is a rich source of information with regard to employee-benefit trends, as well as a number of other things. KFF has state data-fact searches, through which you can find out how many people are insured in a state and what the cost of insurance is in that state, on average, and its demographics.

Pharmetrics (www.pharmetrics.com) has a Web site that you may not have heard about before. The company is a supplier of information to the pharmaceutical industry. The pharmaceutical industry, obviously, is very interested in the cost of medicine and the number of people that have particular diseases. So Pharmetrics created an organization that has obtained data from numerous health plans. They assert that they have 40 million lives worth of data embodied in this data set that they offer.

Milliman USA has explored the Pharmetrics' data as an interested party. As I mentioned, Milliman buys data from a variety of sources. We found that the 40 million lives is kind of elusive, because there is a lot of turnover from year to year. That is, data for these individuals may not necessarily be for the entire number of years that Pharmetrics has collected data. Therefore, it is difficult to use it for longitudinal studies. And they do not always have the amount of CPT detail that you might like to have, or the diagnosis detail. They don't have benefit information available, so you can't tell whether it's an HMO benefit plan or a PPO benefit plan that's at play for a particular set of data that you might be considering. But it does have a lot of information about pharmacy. And it does link the pharmacy data to the medical claims, which can be very valuable.

For those kinds of studies, Pharmetrics can be very powerful. However, running a risk-adjuster program that takes in diagnosis codes and CPT codes to produce risk scores for subsets of data, may not be able to be done that easily. As I said, there are a lot of variables that could be interesting, especially if you want to create pharmacy studies.

There are many tools to which an actuary should have access such as the following manuals—International Classification of Diseases-9th Revision-Clinical Modification (ICD-9-CM), Common Procedural Technology-4 (CPT-4), Current Dental Terminology (CDT-4), Diagnosis-Related Groups (DRG) Handbook. The Dorland Healthcare Information Web site (www.dorlandhealth.com) is a place where you can buy some of these things. The Rand Reports (www.rand.org) has an insurance study that everybody probably should read at some point to understand how they try to scientifically determine what impact changing various levels of coinsurance and co-pay have on benefit usage. It's something that a number of people have

turned to, time and again, to try to get a handle on the impact of benefit differences on the use of health insurance benefits. It is becoming outdated, but it is still one of the best sources for this sort of information.

The American Medical Association (AMA) sells information that you might want to use when you are doing your network development (www.ama-assn.org). Some of this information gets into their expense levels and the impact that malpractice insurance is having on their expenses. We have used it for studies like that in the past.

Open Minds (www.openminds.com) is a behavioral-health periodical that has a consulting group behind it. They have access to a variety of data that they've accumulated from their clients.

The Red Book contains pharmacy data. It gives you information by therapeutic class, and so forth. The average wholesale price is probably the primary thing that Milliman has used it for in the past. The average wholesale price is sort of a fictitious price that continues to be interesting to people that are pricing drug products. You have got to start someplace, I guess. This is one of the places where information is reported.

Periodical examples—I go to Health Affairs (www.healthaffairs.org). Health Services Research (www.hsr.org) is another interesting place for information. *Journal of Health Politics, Policy and Law* (www.jhppl.org) reports policy issues. *Medical Care* (www.LWW-medicalcare.com) reports data. One recent article gave cost by disease. You will find interesting studies in journals like this. I recommend those periodicals for seeing what is available. Aspen (www.aspenpublishers.com) publishes an enormous amount of material. Its *Managed Care Quarterly* can have interesting reports. *The McKinsey Quarterly* (www.mckinseyquarterly.com) addresses high-level strategy issues. They use their think tank to explore various alternatives for the health care system.

Centers for Medicare & Medicaid Services (CMS) gives you access to the *Health Care Financing Review*, which is a periodical that gets into Medicare information in detail (www.cms.hhs.gov). It gives you guidance on hospital costs, and I would suggest that Web site for anybody that's got Medicare + Choice or Medicare Supplement programs. *MedPac* (www.medpac.gov) similarly spends a fair amount of time looking at hospital payment levels, and Medicare's reimbursement levels, comparing them to cost.

We use another Aspen publication from time to time, *Journal of Health-Care Finance*. It has interesting articles, data of a sort, not detailed claim data that we have been talking about, but summary-level information that can be valuable.

The work that actuaries do will become more valuable as more data is made available through the states. I would normally be on the side to not have a lot of money spent by a government to do some of these things, because it does take

money. However, the value of having access to all payer data in single sources is enormous.

MS. DENISE LOVE: I appreciate this opportunity to speak with actuaries who represent a significant audience for state and federal data sets. I spend a lot of my time at NAHDO promoting the development and expansion of statewide health care data, helping states justify their expenditures in these data. Hopefully, I will convince you of the importance of this data, and you can go back to your state and advocate for more health care data to your legislators. It was an actuary that saved the Utah data system telling legislators that the "actuarial profession in Utah would come to a halt without publicly available hospital data."

In addition to connecting you with public health databases, my goal is to introduce you to at least one new portal or one new data set, and convince you that public health data are a good resource. With the publicly available data, you are increasingly getting better documentation and user support.

Publicly available data are generally grouped into the following categories:

Federal Health Surveys and Administrative Data Systems

(<http://aspe.hhs.gov/datacncl/DataDir/index.shtml>)

Federal-state Data Systems:

- Vital records system <http://www.cdc.gov/nchs/nvss.htm>
- Behavioral Risk Factor Surveillance System (BRFSS) <http://www.cdc.gov/brfss/index.htm>
- Healthcare Cost and Utilization Project (HCUP) <http://www.ahrq.gov/data/hcup/>
- Pregnancy Risk Assessment Monitoring System (PRAMS) http://www.cdc.gov/nccdphp/bb_prams/index_longdesc.htm
- Medicaid <http://www.cms.gov/medicaid/>
- www.mentalhealth.org/cmhs/index.htm

State-based systems and links to state-level information:

- Vital records (<http://www.cdc.gov/states.htm>)
- Data with federal component:
 - Cancer registries <http://www.cdc.gov/cancer/npcr/index.htm>
 - HCUP – list of states <http://www.ahrq.gov/data/hcup/hcupsid.htm#head2>
- CMS_data contacts <http://www.aphsa.org/links/statecontacts.asp>
- Minimum Data Set (MDS) (<http://www.aphsa.org/links/statecontacts.asp>)
- Workforce/licensure (<http://www.docboard.org>)
- Financial aggregate reports – from state DOH sites:
 - (<http://www.cdc.gov/other.htm>)
 - <http://www.aap.org/research/medicaid.htm>
 - <http://www.nahdo.org/membersites/asp>

International health data sites that may be of interest to actuaries include:

- United Nations <http://unstats.un.org/unsd/>
- OECD <http://www.oecd.org/EN/home/0,,EN-home-20-nodirectorate-no-no-no-20,00.html>
- Pan American Health Organization
<http://www.paho.org/Selection.asp?SEL=HD&LNG=ENG>

Before you get started "surfing" to cool sites, you might want to check out this tutorial about surfing and how search engines work:

<http://www.nlm.nih.gov/nichsr/usestats/search.htm>

What we are seeing now is the development of public domain analytic tools, such as the Agency for Healthcare Research and Quality's Hospital and Prevention Quality Indicators (www.qualityindicators.ahrq.gov/data/hcup/inpatqi.htm). It used to be, if you wanted to risk-adjust state hospital data, you would have to make an agreement or buy a severity-adjustment system and license agreement, but these public tools are a cost-effective way to explore health databases for screening statistics. I will show you some tools that are in the public domain that you can apply to your own private data sources at no cost. The tools work with SAS or SPSS and are fairly well documented.

With the emerging interest in quality and patient safety, new public resources at the state and federal levels are becoming available. I mentioned the quality indicators from AHRQ, and Richard mentioned the National Quality Forum (www.nqf.org). Medicare is now publishing nursing home report cards on the Web (www.medicare.gov/nhcompare/home.asp), and states are publishing state-level quality statistics:

- www.phc4.org/idb/hpr
- www.thcic.state.tx.us/iqireport2001/iqireport2001.htm

I think that the biggest downside to publicly available data is their relative lack of timeliness. National data sets can be up to two years old. Statewide health data sets can be up to one year old. But mostly the data is used for benchmarks and trend analysis, and hospital statistics don't change that much from year to year, so for these purposes, timeliness is not a huge factor.

Public agencies spend a great deal of time collecting and editing the data, updating errant data as necessary. So, timeliness will always be a challenge with public data.

Collecting identifiable data permits public agencies to link records across data sets, across providers and over time. These linkages are essential to outcomes studies (e.g. readmissions and out-of-hospital deaths), but with HIPAA, it is getting more

difficult to obtain identifiable data. Most of you work in the private sector and are subject to the HIPAA privacy regulations related to data sharing and disclosure. Though public health is exempt, we are seeing restrictions in the public domain as well. I won't go into details, but if you are using public or private data, or sharing these data with others, be sure to understand the terms of obtaining and using the data. You may have to sign a data use agreement. Unless you are conducting research under the auspices of an Institutional Review Board approval, you generally cannot link the data with other sources. Unless you are conducting your analysis under the provisions of treatment, payment or operations, you generally cannot attempt to identify a patient. With the Healthcare Cost and Utilization Project (HCUP), you cannot try to contact the providers. This is just a friendly reminder to those of you use public health data sets, read those data-use agreements, because the federal government will enforce them and make an example of you if you violate the terms.

I will show you some surveys and examples that are maintained by state or federal governments, or both. Administrative data include birth, death, billing, hospital, HCUP and vital records.

One type of public health data is surveillance and registry databases. Surveillance data are the data providers or clinicians are required to report to public health—for instance, lead screening, asthma admissions, injury reports and communicable diseases. Registries are surveillance databases with a lab component. The common uses of these data sets are for epidemiology (calculation of rates, prevalence, and causation).

Trends—you are seeing a lot of the data in public agencies go out onto the World Wide Web. Many of these sites are interactive query systems, permitting you to build your own statistics. The significance to an actuary is that you can "shop" or preview the data set for content. For instance, if you are considering a research project or if you are thinking of purchasing a data set, you can go onto these big sites and look at significance, volume. Is this something that it has variance from year to year? If so, invest in the data set. But if it has only 30 cases, you might just forget it.

Federal interactive query example: www.ahrq.gov/data/hcup/hcupnet.htm

State sites: www.nahdo.org/hidsc/interactive.htm

Again, watch out for data-use agreements, because even on the Web, when you say, "I accept," you better know what you are accepting. Data-use agreements under the HIPAA privacy rule are a regulatory means to control access to the data. A data-use agreement is to be used in these limited data sets that are in the public domain, but there are some stipulations.

DataFerret is an interactive Web site developed by the Census Bureau, and is a good tool for accessing large national surveys (www.ferret.bls.census.gov). You can

create your own tables. You can pick and choose whatever variables you want. You can select which ones you want on the table and which you want as rollover or column variables. It could be used to find significance or research your data set before you invest in it. For serious users, you can get the database and the analytic tools.

Another major source of hospital data and analytic support tools is HCUP (www.ahrq.gov/data/hcup). HCUP is a family of databases and represents a federal-state partnership between the Agency for Healthcare Research and Quality (AHRQ) and up to 33 participating states that provide their statewide hospital discharge data to the AHRQ. The National Inpatient Sample is a 20 percent sample of all HCUP discharges available on CD-ROM or you can query it on the Web. You can also obtain state-level data for that state's entire population in HCUP or standardized format at www.ahcpr.gov/data/hcup/hcupsid.htm#head2. These state inpatient data sets (SIDs) provide state-level, complete data sets in HCUP or common formats so one can conduct multi-state analyses without having to recode each state's data. Analytic tools are available to analyze these data sets, including the AHRQ quality-indicator software.

How would an actuary use these data? Say you work in a health plan and have access to your own proprietary data, including inpatient data in UB92 or billing format. You could download the AHRQ quality indicators and generate your plan's quality, prevention or patient safety statistics by patient and hospital characteristics from the AHRQ QI software. Next, you will want to have context for your statistics. Are the rates in line with other comparable populations or types of hospitals? You can obtain a nationwide inpatient sample and/or several state-level databases and generate standard statistics by patient, geographic and hospital characteristics for national, regional or state benchmarks and compare your proprietary data with these. Though these are mainly screening statistics, this exercise permits actuaries to evaluate quality performance, utilization patterns and variance and begin the process of setting targets for improving care and quality. Actuaries are important power users of this type of data set, and I would welcome any feedback from those who are using these tools and data sets.

So how do you find public health statistics on the Web? The Department of Health and Human Services (DHHS) does not have a central statistical agency. The surveys and the data collection systems will vary in the health part of the government.

I found a gateway to federal data, and it is sort of a mini-portal. It is where all of the surveys and all of the data reside through the federal government, <http://aspe.hhs.gov/datacncl/>. It is organized by the various agencies, and you can go into each of those. I would recommend, if you were not familiar with the federal systems, to use that gateway. It's a clearinghouse of sorts to national health data sets.

Through this site you can access important types of public health data sets. You can go into the federal site of the vital records system. For example, you can go into the Behavioral Risk Factor Surveillance System Web site (www.cdc.gov/brfss/index.htm). The Pregnancy Risk Assessment Monitoring System (PRAMS) is a new surveillance system. It covers 20 to 27 states. There is a link through CDC (www.cdc.gov/nccdphp/bb_prams/index_longdesc.htm). The organization samples a certain number of births, then interviews the mothers about risk factors that she may have been exposed to during her pregnancy.

There is no gateway for state data. States vary in their organizational structures, regulations, data access. But you can get to the states through some federal "mini portals," depending on the data set. One of them is vital records. The CDC maintains links to state vital records agencies that general local birth and death statistics. Medicaid lists all of the state Medicaid contacts on the Web (www.cms.gov). Again, you can find HCUP state data at www.ahrq.gov/data/hcup. I think the documentation is good.

If I wanted to access the minimum data set for nursing homes, I likely would have to go directly to the state Web sites and locate the proper state agency that maintains the data set (www.cms.gov/medicaid). States also maintain health workforce and licensure data, but your best bet for locating which states maintain these and how to access these data is at www.docboard.org. This site has a list of all of the licensure agencies in the state—licensed physicians and professionals. And you can search the data on the Web. The problem with state Web sites is that they are not all organized in the same way. And you can spend a lot of time trying to figure out where to go for data, but there is no super catalog of these data, so I look at portals as being a replacement of that.

I have shown you some mini-portals into certain data worlds. Search engines are another way to work within those portals. You can use GOOGLE, but if you use a search engine in a portal, it is going to be a lot more sensitive. Searching directly through a portal will help you find what you need a little quicker.

There are other portals I have used and should mention, though many of you likely are aware of these portals. The statistical universe used to be at the University of Michigan (<http://www.lib.umich.edu/govdocs/sthealth.htm>), but now can also be found at <http://web.lexis-nexis.com/statuniv>. This site has a huge list of state government sites and census information. You can do a search, site map, life tables, IRS, mental health, etc.

You can also find the Web equivalent of the printed American Statistics Index, an index to international statistics loaded in the document center. The site contains all of the abstract statistical publications and approximately 15 percent of federal statistical publications. You can get to the data on nursing homes, risk behavior, etc. They have tutorials. It's a fabulous site.

The National Library of Medicine has a super Web site (<http://www.nlm.nih.gov/nichsr/hsrsites.html>). It is the gateway to text, health resources, search engines and databases. You can even get health-services research projects in progress, unpublished. I went to the NLM site and queried cancer data:

- <http://www.nlm.nih.gov/medlineplus/healthstatistics.html>
- ADD "CANCER DATA":
- <http://www.cdc.gov/cancer/npcr/uscs/pressrelease.htm>

I got some links and news articles. Since I wanted cancer data, I got into the CDC National Cancer Data gateway. It has whatever data you want—from skin to colorectal and special populations. Then I wanted state information. The Web site breaks it down by the state statistics. What other cancer resources are out there?

- National cancer data:
 - <http://www.cdc.gov/cancer/natlancerdata.htm>
- State cancer data:
 - <http://www.cdc.gov/cancer/dbdata.htm>
- North American Association of Central Cancer Registries (NAACCR)
 - <http://www.naacr.org/>
- NCI's SEER program: <http://seer.cancer.gov/>
- NCI CancerNet Web site: <http://cancer.gov/cancerinformation>

This all came from this NLM search. It is just one example of searching through the NLM portal and funneling down to the topic that I was interested in.

In 1998, my organization, NAHDO, created a site that was customized for health actuaries. It took so long to create it that the federal government started doing it for us. It is still out there, but it's not maintained very well right now. It has a search engine and various catalogs and categories, such as aging. You can add sites that you think are useful (www.nahdo.org/soa1.html). We are not now maintaining this site, but it might provide private-sector links to augment the federal site portals for some of you.

An HIPAA site NAHDO developed for states providing general information about standards and HIPAA is (www.nahdo.org/project/index.asp). It is a nice tutorial if you do not know about HIPAA and want to know the process for adopting standards and finding the standard that fits when you are designing. These are for state architects who are designing a data-collection program.

If you do not know about the program, how do you change a data element? This site tells you about the Data Standard Maintenance Organization, and I encourage any of you, when you find out who in your organization tracks standards, make sure that they attend its sessions, because those who show up decide what the standards are. If you don't show up, you don't have a say.

MR. KIPP: I forgot to mention Solucient Publications (www.solucient.com/publications/pubs.shtml). Solucient's Web site is probably in the top four or five sites. The company sells large data sets. It bought all of the inpatient discharge data. It has gotten data from some of their clients. It is selling decision-support tools and trading for data, and it has accumulated a good-sized data set that you can buy access to. Just be prepared to come with your checkbook. It's a fairly expensive proposition.

MS. LOVE: The nice part about the publicly available data is that it's usually cheap and accessible. I would encourage you to visit www.ahrq.gov/data/hcup for a suite of data and tools. You can download the quality-indicator software that takes your administrative data, hospital data, and actually runs through 33 quality indicators, prevention indicators, things like asthma admissions to hospital, and you can download and run the patient safety indicators, which are brand new. The data formats for the nation and the state are in uniform formats. You can acquire any number of state data sets in a common format, plus a national inpatient sample, and run it internally with your own administrative data. I think that would make a powerful case for what is happening around variance and outcomes in your organizations. The software is free, but it can make a powerful argument. Stanford University developed the software and validated those tools.

I want to acknowledge the following individuals who have shared their expertise and work with me over the past few years:

- "Health Statistics: Finding and Using Them," Dan Melnick Research <http://www.melnickresearch.net/sys-tmpl/door/> Practical Resources for Public Health Data Organizations, and
- Marjorie Cahn, head, National Information Center of Health Services Research and Health Care Technology (NICHSR) National Library of Medicine, NIH, HHS. nichsr@nlm.nih.gov, www.nlm.nih.gov/nichsr