

# RECORD, Volume 28, No. 2\*

---

San Francisco Spring Meeting

June 24–26, 2002

## Session 15TS

### Public Databases and Other Resources for Health Actuaries

**Track:** Health

**Moderator:** GRADY C. CATTERALL

**Instructors:** GRADY C. CATTERALL  
DENISE LOVE†  
TIMOTHY K. ROBINSON

*Summary: This teaching session provides a summary of publicly available data and information sources for health actuaries. Several of the more useful resources are reviewed in detail. At the conclusion of this session, participants gain an awareness of the wide variety of data sources available to practicing health actuaries and the potential usefulness of these sources.*

**MR. GRADY C. CATTERALL:** This is supposed to be a teaching session, but it is about six times the size of your optimal small-group teaching session. We'll try to answer any questions that come up and give you as much hands-on and detailed information as we can.

I just joined The Lewin Group as a senior manager in the last week. My major area of practice is Medicaid managed care plans. I'm also the chairman of the SOA/NAHDO (National Association of Health Data Organizations) Work Group, which is a subset of the Joint SOA/Academy Committee for Communication on Health Issues (JCCHI). The work group's purpose is to find and review ways to update and revise the SOA's Health Web site search module, which can be found on the NAHDO website. I'll be talking more about that later on.

Our first speaker is Tim Robinson. He's a consulting actuary with the NiiS/Apex Group. They do consulting for health plans, plan sponsors, reinsurers and managing

---

\* Copyright © 2003, Society of Actuaries

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

general underwriters (MGUs). Tim's work involves a combination of working with primary health plans and with stop loss plans. Prior to that, he worked for Avondale on catastrophic care carve-out plans. Over the last year, Tim has been responsible for updating and maintaining the Health Resource List on the Health Section page of the SOA website. No one had been maintaining the list for a while, and it needed a lot of updating. I think Tim has done an excellent job getting that up-to-date and keeping it that way.

Next, we have Denise Love, who is executive director of NAHDO. NAHDO is a non-profit membership and educational organization dedicated to strengthening the nation's health information systems. As executive director, Denise is actively involved in national standards forums, developing integrated Internet systems and establishing analytic frameworks for major health data sets. Prior to becoming executive director of NAHDO, Denise was director of the Office of Health Data Analysis for the Utah Department of Health. She's also an adjunct faculty member in the Department of Family and Preventive Medicine at the University of Utah School of Medicine. We will start with Tim's presentation focusing on the Health Resource List on the SOA Web page.

**MR. TIMOTHY ROBINSON:** As Grady mentioned, for the last year or so I've been responsible for keeping the SOA/Academy Health Resource List up-to-date. Before I start, I'd like to get a sense of how much work we need to do or maybe how much better promotion we need to do. How many people are familiar with the Health Resource List on the Health Section Web site? Okay. That's not too good of a sign. My next question, is how many people have actually used it? That's an even smaller number. We'll try to point out some of the benefits and some of the reasons why we tried to create this tool for health actuaries to use.

These are just some of the topics, and I'll go through these fairly quickly because there's actually not that much to cover in-depth. It's more a process of making people aware of what's there and what some of the tools are, as well as some of the links. Basically, as I said, what we've done is compile a list of current articles and reports and have tried to keep this an exclusive list of items that are going to be of interest, we think, to health actuaries. We only take things from actuarial publications and organizations. Later on in the presentation, I have a list of what some of those are. It's easy to find if you look on the Web site and see what's there.

We've tried to keep a fine line between having a set of information that's so broad that it isn't as useful because it takes you so long to find what you want, versus making it too narrow so that it doesn't meet the need of providing enough information. We've tried to have a place where health actuaries can look for current topics of interest that are being discussed in the industry. That's the reason why we really focused on looking at actuarial publications in order to put them on this specific list. A lot of the information, to go beyond that, gets into some of the things I think Grady's going to talk about when he discusses the NAHDO search

engine and some of those other resources. What do we mean by current, as far as current articles and reports? It varies a little bit at this time. I think we're still trying to figure out what the optimal time period is to keep. For the most part, for periodicals, we've only kept things on the list back to the year 2000. There are certain reports and other things that really haven't been updated yet, or that seem to be of more relevance. It's kind of a judgment call, but some of those have been left on there. It's a little bit of a mixed bag right now. We've tried to not perpetually just add and add to it without looking at what's still relevant or what's still timely.

The easiest way to find the list, if you're fairly familiar with the SOA Web site ([www.soa.org](http://www.soa.org)), is to go to the "Special Interest Section" link and then to the "Health Section" link under that. If you scroll down under the Health Section page, you'll find a variety of things, and one of them says SOA/Academy Health Resource List, which gets to all the information that we'll go through. You can also click on a link to "Health-Related Materials Produced by the Academy," and that actually takes you to the Academy's Web site. It takes you to a page where there's a well laid-out list of various Academy publications sorted by topic. In some ways it's organized quite a bit better than the SOA part of it at this point because it's actually sorted by topic, so it's easier to go through and search for specific things that you're interested in. We wanted to include that link in the SOA Health Resource List as well just to make it more comprehensive without duplicating all the items on the list. If you're in the SOA Web page looking at some of these items, just click, and you can go directly to the Academy list as well. If you're on the SOA Web page, you can find it pretty easily, actually maybe not easily enough because only a few people appear to have used it, but hopefully that'll change if we can get the word out.

There are many reasons or benefits that we thought of in putting a resource list like this together that consists specifically of information of relevance to actuaries. One reason is that there's probably a pretty typical pattern of getting so much information in your in-box. You get the various publications and reports, and a lot of times they pile up. You might want to scan the list without having to actually go through and read through every periodical in order to find the articles of interest. By the time that happens, they're not really current anymore. As those periodicals are published and the newer versions come out, we've tried to scan them and take the articles that are specifically related to health insurance. It's a little bit of a judgment call as far as who's to say it is of relevance or not, but we're trying to take the material that is clearly targeted towards that audience, link it and update it on a periodic basis so that that resource is available.

Another benefit that I found in getting involved in this process is that it does expose you to some of the periodicals and reports that you may not typically look at if you're only focused on reading the *Health Section News* or *The Actuary* every month. We've also included articles and reports from some of the other section newsletters. So, you may not find yourself always looking at or subscribing to the Smaller Insurance Company newsletter or the Computer Science newsletter, but if

there's an article that appears that may be specifically related to a health insurance topic, we've tried to pull those in as well. Rather than having to look at all those documents that you may not even be aware of, it makes it easier to find things in one place. Again, the point is that we've tried to make everything on the list relevant to health actuaries, so, at this point, we focused on just the actuarial publications rather than broadening it.

Another feature that's pretty useful are the direct links to some of the articles and reports that are mentioned on the list. Rather than having to go through the typical search engine process, if you find an article, say, in the *Health Section News*, you can click on that, and it'll go directly to that article.

This is just a list of the broader categories of periodicals that are included—the SOA-sponsored publications—and you can find this if you just look on the resource list. Again, this is an evolving thing. As Grady mentioned, it was probably within the last year or so that, through the Joint Communications Committee, we've gone back and tried to update this in terms of removing things that are either outdated or irrelevant and we also add things. That includes updating articles as well as adding some periodicals or some different types of reports that weren't included before. It's an evolving process, and any feedback on what items should be added or maybe taken off would be really helpful to us.

Similarly, for the Academy piece, when you look at the "Health-Related Materials Published by the Academy," the items that are listed by topic. It'll say "Medicare Issues," "Long-Term Care," "Genetic Testing" and things like that, and list all the various committee and work group reports, monographs, issue briefs, etc. At this point, I don't think the *Contingencies* articles are linked, and I guess that's something that we might want to talk about internally with the communications committees if that's one of the enhancements we might want to make to the resource list. *Contingencies* is obviously a major publication that's not included on the list yet. Just to list some categories and maybe some of the types of resources that we haven't included at this point basically gets back to trying to keep the list fairly well restricted so that we can be pretty sure things are going to be of relevance to health actuaries. Obviously, we haven't tried to make this a resource that would encompass non-actuarial publications, even though there are obviously many industry reports and periodicals that people use in their work. We didn't want to broaden the scope too much. We haven't linked to any of the proceedings from meetings such as this. It seems like that's something that's fairly easy to do if you're interested. If you look back at the last several years of spring meetings or annual meetings, you can see what all the different sessions have been related to health insurance. That's another good way to search for current topics, and it seemed redundant to list all those, at least at this point, on this resource.

We arbitrarily defined what was current and what wasn't. But, just to have a cut-off point, as part of the update process, we took off most of the articles dated prior to 2000. I think you can still find all those articles if you go through the usual library

search process, but we decided not to link them onto this resource list because we wanted to keep it a little bit more up-to-date in terms of timing. We've tried to exclude topics not primarily related to health insurance, and again, that could be considered a judgment call. If we look at the table of contents of the Reinsurance Section newsletter and try to take things out that are of specific interest to health insurance actuaries, it would depend on what people's interests are.

This is the process we've tried to go through to update the list over the last year. We probably haven't quite done it quarterly, but that's been our goal. It's a bit of a challenge because things don't always come out consistently. The different section newsletters come out at different times throughout the year. So, usually about three or four times a year I'll try to assemble everything that's come out since the last update, and a few of us on the Joint Communications Committee will go through that table of contents, look over the articles and try to figure out which ones should be included and which ones should be eliminated. It's an ongoing process.

We welcome any suggestions that people have as to what types of different reports or periodicals we might add that aren't currently included, or what items are on there that people don't find of relevance anymore. These are just some of the examples I thought of in putting the presentation together. There are a few of the seminars from the mid-1990s. The proceedings were linked on there, and we haven't taken those off yet, but that's probably another place that we can update. We can add different types of seminars, some of the more recent ones.

I also think the ability to sort by topic would make it more useful, and that's something that the Academy piece of this currently does. It makes it a little less user-friendly on the SOA piece because you have to look at each periodical and know what you're looking for in terms of topic, but it does give you a comprehensive listing of what's currently being discussed. It's just that it's not quite organized in the same way. That's another piece you might be able to look at in the future. If you do get a chance to look at the list and have any suggestions about improvements we could make, that would be great.

**MR. CATTERALL:** I forgot to mention that Denise was the only panelist for this session last year, and it was a very well attended and well regarded session. We're very happy to have her back for this year's session on public databases.

**MS. DENISE LOVE:** I'm acutely aware that I may be the only non-actuary in the room, and I'm a little intimidated until I think about the fact that this is probably one of the few groups I speak to who, like my own group, doesn't think of data as a four-letter word—you really are into it. So, with that said, I'm hoping to show you today some of the projects that not only my organization, NAHDO, is working on, but also our partners in the federal government and state. The members of NAHDO are stewards or owners or developers of big data sets at the state and federal level, and many of those agencies have been fabulously successful over the many years

of collecting data from hospitals or different health entities or surveys. But we haven't been very good at getting it out to beyond the agencies or their immediate audiences.

What I want you to take away from here is that the Internet has changed everything, and even the federal government has made a very large commitment to developing ways that broader audiences can access the tools. I may not do justice to my federal friends, but I'm hoping to at least engage you, if you have not already been engaged, with some of the new things coming down the pike for accessing publicly available data. I'll go through a few of those, and, in fact, as recently as this morning, someone sent me another site that I was not aware of that I added. I'm going to deal mostly with health databases, but how do I characterize public databases? You can say federal, you can say state, you can say private, you can slice and dice it different ways, but I'm looking at the data catalogs. You have your big federal surveys, some of them longitudinal, some of them not; you have your national surveys; and you have your Centers for Disease Control (CDC) surveillance or your reports or your morbidity reports. Then you have population-based data, and I'll talk a little bit about the Healthcare Cost and Utilization Project (HCUP), something near and dear to my heart because, I'm a subcontractor to that. All of you are aware of census data, but then you have state datasets and health datasets that are evolving or have been around for some time.

I also thought about ease of use and access. I think the surveys are fairly easy to get access to or download on the Web, and most of you or many of you have taken the National Hospital Ambulatory Medical Care Survey, for instance, or a sampling of hospitals, or the Health Interview Surveys. You can usually get a version downloaded from the Web, though timeliness is always, for reasons we can talk about later, and reasons I can't change in the short run, an issue with federal databases. You also have your CDC statistics. I put in the mid-level MEDPAR, vital records, some of which are fairly easily gotten. But the difficulty for my agency, for my members and everyone else lies in getting micro data directly from some of the Medicaid agencies such as the Centers for Medicare and Medicaid Services (CMS) that have patient characteristics and are identifiable. Those are usually protected, and we can talk about ways to access them, though it is not easy for legal and confidentiality reasons. But there are ways to get at some of the data that I will discuss later.

I'd like to talk about state-based datasets where I do most of my work. Hospital discharge data is one example of a population-based dataset that I know a lot about. Again, everyone here knows about the vital events, the state surveys, the behavioral risk factor survey and health status surveys. Most states have that data at their state level and some of the national CDC compilations of the Behavioral Risk Factor Survey System (BRFSS). We can go back and talk about those, but let me just do a rundown because I think any discussion of publicly available data needs to address the fact that how we get that data has changed. It used to be that

we knew where our little pockets of data were, but I think that the Department of Health and Human Services (HHS) is doing a pretty good job of changing that.

The first tool I wanted to talk about today is one that some of you might not know about, because I didn't know about this until I was recently appointed to a national academy's panel on race ethnicity. I wanted to know what was happening with federal databases and collection of race ethnicity in health databases, and I came upon the directory of HHS Data Resources. In essence, it is a compilation of all the major datasets that HHS maintains. Now, that doesn't necessarily mean special project databases or one-time collections. These are the ongoing databases that HHS directs and maintains. This catalog was developed under the direction of the HHS Data Council, which is a council that represents all the major agencies in HHS at the federal level. They meet regularly, and they attempt to look across the whole agency to see what's going on and try to make things more uniform, rational or cogent as far as their data efforts go. This directory expands the 1995 HHS Directory of Minority HHS Data. Apparently there was this thing developed in HHS that now has been updated and expanded, and it includes published aggregate and public use micro data, including the recurring surveys, disease registries, and, again, some one-time studies that have broad interest. I don't know if it's the council or otherwise that defines it or how they do it.

For example, I went into the HHS directory and looked up Health Care Financing Administration (HCFA), which we know now as CMS. This page has the enrollment database, health insurance skeleton, eligibility, write-off, HISQ file, national claims history, etc. It lists and catalogs their datasets online, and then when I clicked into the Medicaid Statistical Information System (MSIS) Personal Summary File, it told me this is the template that all of those databases follow.

It's not actual data, but for each dataset, the acronym tells you which agency maintains it, the description of that dataset, the race ethnicity, how it's collected and in what format in that particular database. It does describe the data limitation, the status, the years, how to access the data, the Web site and other contact information. I think that if you're going to start perusing and using databases, the Web site for this directory is very helpful, and, as far as I understand, they will continue to maintain this. One bad thing about the Web is that once you update something then go away, it's irrelevant in a year. Hopefully, and from what I understand, there's a commitment to continue maintaining this data. That's one tool I wanted to share with you today.

The National Library of Medicine has a fabulous number of resources. I won't go into each one. This database contains information about research datasets, instruments, indices used in health services research and behavioral and social sciences with links to PubMed and additional resources. The Web site is [www.nlm.nih.gov/nichsr/hsrr\\_search](http://www.nlm.nih.gov/nichsr/hsrr_search). It cites the datasets by title.

These are a few examples from the dataset:

- City's Mortality Reporting System
- Public Use Micro Data Sample for the Older Population
- Longitudinal Follow-Up to the National Maternal and Infant Health Survey
- Blood Level
- Epidemiology
- Surveillance Program"
- Trauma Registry
- Work-Related Injury Surveillance–(I think that's Alaska)
- Physician Master File
- American Indian- and Alaska-linked Birth/Death Record Database"

These are specialty datasets by title, and then they go into another family of links that are instruments and indices by title.

- Ambulatory Care
- Medical Audit Instrument
- Referrals to Specialists
- Health Profile
- DUKE
- Severity of Illness Checklist
- General Health Questionnaire
- Opinion Survey

These are special indices and instruments that have been developed under the Health Services Research rubric for power users of the data. Then it lists software by title, explaining Filemaker Pro, Explorer, SPSS and some of the tools that researchers might use. I must be honest, I did not go into this part, but I wanted to share this with you.

The next tool that I'd like to share with you is the Data Web ([www.theDataWeb.org](http://www.theDataWeb.org)). The Data Web is something I think you should also check out. The Data Web is a joint project between the U.S. Census Bureau and the CDC. My friend from CDC sent me some of his slides. I'm not going to pretend to know everything that was in his mind when he put the concept together, but Data Web really is a distributed information system and infrastructure, whereas the state data and the agency data reside where they reside. It's not centralized in some grand database in the sky. It's a distributed system that has client interfaces where you can access the data through a user-friendly interface.

Again, he scaled the data as pretty aggregate data for public use—easier to get, and easier to use for people like you who will be more research or- decision-support-oriented. You need more raw data. You need more power of the data. And the data needs to be more sensitive. So, again, you could call each agency, try to build your own analytic files and repository, but this is something that, at least for first cuts of data, you can play with. I'll talk about it a little more, see what that



data is like or if you even want to get more data, so you don't have to invest the time and energy to go out and build these repositories just to find out that one dataset is worthless for you. You might be able to test-drive it.

Again, what is the Data Web? It is a large repository of needed data. I think that I could argue: "Needed by whom?" Feedback would be important if you don't see what you need on it. It's a mechanism to access data across the Internet at any participating site regardless of your platform. It consists of a uniform data dictionary and metadata file that document the data for human use. It gives you information to use in your client applications. Again, they're services that support social science applications, including time series, survey handling, data transformations and file formats that most desktop publications support.

The client system would be the person requesting the data. They go to the metadata service and find out what is cataloged, how it is defined and what the fields are. When you submit a request, it comes through and transforms the data. You can recode the data back to your system. You can reaggregate the age groups if you don't like the age groups that the agency has defined. It has some powerful transformation abilities. The person that gave me the slides is the actual developer. But, again, the idea is it's distributed, and it accesses the data that the agencies have prepared for the Data Web, and NAHDO is one of the sites. So, I know a little bit about the preparation of the data.

One of the applications that I've used, and I'll show you some examples, is called the Data FERRETT. You can do some exploratory analysis. And, again, it's not for your casual consumer user. It is for a more sophisticated user. Some time ago, I built a query concerning poverty health insurance by state. The datasets are current population survey, folder, basic and the March 1999 supplement. Again, you get the variables listed that are cataloged in the metadata file. You can use the control key to select what field you want out of that particular database.

Then you browse you're the variables, selecting them all or selecting specific variables. Your variables are listed, but the neat part is you can drag them. So, it has some user-friendly aspects, and you can drag them to be column variables or row variables. You can switch them around and play around with what you want. Then you hit the query. They have hot reports, which is another feature that they're developing and enhancing. You can get a quick graphic output or just a tabular output of whatever cross-tabs you check from the variables there. You can also integrate some of the datasets. You don't have to just work with one dataset at a time. You can integrate some of the fields and datasets and build your cross-tab on multiple datasets. This is another tool you can use to test or explore the utility of datasets you might not have previously used.

Now, NAHDO has taken emergency department datasets from three states—Maine, Hawaii and Utah. We loaded both the micro data and the aggregate for Maine on the Data Web. We had to develop a new metadata dictionary and help Census and

CDC adjust and rewrite some of the applets that allowed Maine's data to be queried on this. I then had to embargo the site until I got the blessing of the State of Maine and their hospitals. We embargoed it on a login site, and we're reaching out to other states. This is very tentative, but we're looking at the birth and death files of California Vital Records to be put in a special site on the Data Web. I'm not convinced that it's the tool that states want to use, for many reasons I don't need to go in here, but my thinking is if people become accustomed to or like using the Census or the FERRET for other datasets, it's a common interface and a similar tool where they could access some of the state data without having to shift gears too drastically.

This is quite preliminary, but I think it's another way to open up state data to users. Now that we've got it down, we think we might be able to do similar datasets. It took six months to load Maine's, but we think now we could probably do other state emergency department datasets in a month or so. I think we've figured that one out. The beauty of it is it's running on the NAHDO server. It's not running at the Census site. So, in a sense, states or data users could think that they still own their data. They're not sending it off to Washington, D.C. What you're doing is formatting it in a uniform way, and the user doesn't know where the data really resides. That's the beauty of the Data Web, and I hope you have a chance to play with it. Again, they're working on the hot reports, dynamic tables, texts and graphics. There's some sort of Web directory coming out, but it's not out yet. So, watch what's happening with the federal government because, as the Internet evolves, you'll see more sophisticated access.

One area that CDC is targeting is what NAHDO's project—emergency department data at the state level. But they have all major surveys up, including the ambulatory care surveys, the risk factor surveys and the youth risk factor surveys. I think that this would be a useful tool for an actuary going through his daily work. That's my assumption, but, again, I'm a non-actuary, so correct me if I'm wrong.

Again, some other data objectives include the 2002 Current Population Survey (CPS), County Level Poverty Estimates, Economic Census, and the FBI Uniform Crime Data by County, which should be interesting. Now, for those of you who want to go right to this site and use it, there are some applets you have to download, and you have to read a data-use agreement. You just can't jump right on the site. You have to login and download the applets, but once you're on you can go back to the site on fairly easily.

Let's shift gears again. There is another system that I'm pleased to tell you about today. How many of you have used HCUP data? I hope after this session we have more hands go up next year. HCUP is really a multi-state administrative population-based data system that includes all payer hospital data in uniform format, primarily formulated for health services research but by no means limited to that. It's a family of products. The key players in HCUP are the Agency for Healthcare Research and Quality (AHRQ), the state data organizations, many of whom are NAHDO's

members that maintain statewide hospital discharge data at the state level with subcontractors including MedStat, NAHDO and Social and Scientific Systems, and the American Hospital Association (AHA), which provides survey data.

It began in 1988 with nine states, and in the year 2000 there were 29 states. It is a growing enterprise. Some states are providing in-patient data only. All of them provide at least in-patient data, but some states provide only in-patient data. Some states are providing both in-patient and ambulatory surgery data. Some states are providing not only in-patient and ambulatory surgery data, but emergency department data as well. The other states are non-participating.

There are state in-patient databases in uniform format. There's a nationwide in-patient sample and the ambulatory surgery and emergency department sets. It also contains research reports and statistics. I think I have a link to HCUP Net, and this is another query system that you can reach online. Let's say you're working with some in-patient data, and you want to validate a statistic, but you don't want to buy a whole dataset to have a reference check of the statistic. You can query by patient characteristics or hospital characteristics, one of the cross-tab queries, and see if it's ballpark with yours. That's the minimum use of this, to see if you're in the ballpark with your particular data system, or you can play around with the data to see if maybe you want to order the whole dataset. It gives you a test drive of that dataset. The tools are, I think, the most exciting part of the whole thing, and we'll talk about this more tomorrow at 10:30 with the Stanford team who is taking the dataset and doing a validation of the quality indicators. They're developing clinical classification software. It's not just diagnosis-related group (DRG), major diagnostic category (MDC) grouping or proprietary system you're stuck with, but there are clinical classification software tools that you can download from the Web to group your data into, and you also have quality indicators.

What are state in-patient datasets? They're all hospitals provided by the state. The state gives up all of the data, all discharges and all payers. Then MedStat and AHRQ uniformly code the data and edit it. The good news is that it's edited into a common format. The bad news is it might be the lowest common denominator because some of the states are collecting data a little differently than others. But the state in-patient datasets represent 80 percent of the U.S. discharges. What is the discharge data level of files? There are minimum core data elements in each HCUP that are commonly available across all states. There are some state-specific data elements. Some states collect race ethnicity. Some states collect unique identifiers that can be encrypted. They collect more data, present on admission indicators. The SID will retain those unique data elements, putting the sensitive ones in a separate file not available for public use. The hospital-level files include hospital identifiers and characteristics from the AHA files.

Patient demographics, patient zip code, diagnosis and procedures, expected payer, admission patient disposition are all on the SID, with all of their flaws. A state-

specific SID like Texas or California that collects race ethnicity will have that information. A state that does not collect it will not have it. Severity of illness indicators, if a state indeed collects those, will be on the SID along with birth weight, time of onset of diagnosis and physician identifier. Some states do require more reporting out of hospitals than others. In terms of how it's coded, they've tried to retain the International Classification of Diseases-9<sup>th</sup> Revision (ICD-9) coding in the original form, but the identifiers are, of course, encrypted into synthetic values. The sex, race, payer and other demographics are either calculated or recoded to uniform coding and the DRGs are assigned using external algorithms. In order to include some variables about hospitals (if you're doing some comparative reports or you're adjusting for those factors), they add the AHA annual survey and Medicare public use files from the cost reports. The county codes are from the area resource files, and the hospital zip codes are from the zip code files.

The National Inpatient Sample (NIS) is a pretty cheap dataset, and you get it in a CD format. It's a sampling of community hospitals from the SID, but it's a sampling of hospitals that are representative and weighted for that state's population and experience. It's not a sampling from the hospitals. They sample the hospitals, but it's 100 percent of the hospital visits from that particular hospital representing seven million discharges, and, again, we talked about the data elements. It's stratified, a 20-percent national sample, by these hospital characteristics. AHRQ is building specialized databases. Say you want to do a pediatric study. Before, you had to get the SID. You'd just filter out your pediatric, zero to 18. Now they're building a specialized kids database, with 1.9 million unweighted discharge records, but to represent 6.7 million weighted discharges for pediatric care.

Ambulatory surgery is an emerging dataset. We have 25 states collecting ambulatory surgery data, but not all of them report to HCUP. Again, they are consolidating 15 states and search files. We're working hard to get states to collect emergency department data. Right now they're in a pilot of seven states and learning about that dataset, and NAHDO counts about 17 states collecting emergency department data.

**FROM THE FLOOR:** You mentioned this one dataset is pretty cheap. What is that in terms of cost?

**MS. LOVE:** Yes. I think the NIS is \$150, but it's cheap. It's amazingly cheap, and it's well documented. It comes with a CD packet. It's a nice little binder. It has the documentation and the data elements, and it's quite reasonable. I think it's the best value in data today. It's accessible, and it's a bargain if you're doing national studies. Now, if you want to drill down in a state, that's a little different. If you want to get a SID or a state in-patient dataset or the state's dataset in the HCUP format, states charge anything from \$500 to \$3,000 or more for that particular dataset with all of the discharges. So, it's a little different ballgame when you want a complete state picture.

For all HCUP data products—the microdata and access to the microdata—that you get when you buy the dataset (\$148 or \$150), you must sign a data use agreement. The data use agreement basically says that you cannot attempt to identify any individual by imputation or otherwise, and you can't attempt to identify institutions. You can't make recommendations publicly that this institution is better or is a better performer than others, and you cannot re-release it without the permission of AHRQ. They do have quite hefty penalties. For those that want to do a hospital-level study, the NIS may not be the dataset. You may have to go directly to that state that doesn't have those constraints on the data because most states that do collect the data permit the identity of hospitals, with exceptions.

For HCUP Net, the software is SPSS load programs format so you can use some of the tools that are downloadable on the site. HCUP Net can be queried by patient and hospital characteristics. It's off the [www.ahrq.gov](http://www.ahrq.gov) site. You go to HCUP or HCUP Net, where you can find instant tables, national statistics, the children's hospital stays-only trends and state statistics. They'll link you right into some of the state queries. You might be able to get some basic queries directly from the states participating. And, again, you can calculate some quality indicators. It's pretty quick and it is interactive, but again, it's limited in that you can't really drill down and find out what's going on. You can get some instant stats. It's valid for 1980 through 2002. It's updated nationally and available on the HCUP Web site.

The Clinical Classification Software (CCS), available off the AHRQ Web site, is also available for download for use on your own datasets. This is just an example of a CCS query. It's a single level—tuberculosis, septicemia, bacterial infections—you get the mean charges, and mean length of stay. Then you can get the multilevel breakouts of the tuberculosis, septicemia, and the different kinds of streptococcal, staphylococcal and their charges and their breakdown. You can get the actual software to group your own data.

CCS is an analytic tool that can be used for looking at comparative performance of hospitals by their resource consumption, utilization profiles, conditions by groupings, and then being able to predict future resource use. They have some citations to literature of who's done that and how they're published. Again, the quality indicators are a screening tool. They're not true quality measures. They're used to screen for proxy measures of quality. They're designed for use with administrative data. They're used by hospitals for internal quality improvement, but, again, states are using them as well for some of their quality studies and quality comparative reports.

Some of the research applications and potentials of the HCUP data include access to care inference, impact of health policy changes and small area variation. We talked about how it captures the state data and all of the stays in the hospital. It is a sizable database. You can look at state and sub-state focus, and some states permit subpopulation focus and look at rare disease and procedures because of the number of observations. Again, it includes the uninsured. It's just not a payer

database. It permits trend analysis because it has over a decade of data, and you can look between states and link to other databases, but you can't link it to other databases. That's against the data use agreement.

Again, timeliness for government databases is always a challenge. When you're putting 29 state datasets together, and in a state, with 400 or 435 hospitals, such as Texas, you can only move as fast as the slowest state or the slowest hospital. Then government databases are incumbent, meaning they have an added responsibility that they have to be fair. When the data becomes public or used by the public, the state or the federal agency has to do the due diligence to be sure that the data is as validated as possible, given all the problems with data. That state or that federal agency has to do a lot of back-and-forth updating or work with the data supplier. A government database will never be quick and dirty and probably shouldn't be quick and dirty in my opinion, because it needs to be a reference database, and the care needs to be taken that it's as good as it can be, as imperfect as it is.

There are some limitations of HCUP. We can talk all day about the limitations of hospital discharge data as administrative data. It has power of numbers. It's easy to get. It's fairly uniform. But it has limitations. We don't have clinical variables in that. Patient identifiers aren't uniform. So, you really can't look at episodes. If you are in a proprietary database, you have the luxury of looking across patients across time. There is no revenue data. But there is a charge-to-cost methodology developed by AHRQ, and it's available on their site. I think that's a reasonable approach to imputing the revenue data. Then again, states are just now drilling down or trying to add physician IDs, and it'll be many years before we have a physician ID data added routinely with hospital discharge data.

Table 1 lists some recent studies that the HCUP data has been used for. Some have actuarial interest and some may not. I just thought that it would be helpful to see the links of who's using it for what.

Table 1

## Recent Studies Using SID

---

- Use of major therapeutic procedures: Are Hispanics treated differently than non-Hispanic whites?  
**Andrews & Elixhauser.** *Ethnicity & Disease.* 2000.
  
- The marginal benefits of invasive treatments for acute myocardial infarction: Does insurance coverage matter?  
Brooks, McClellan & **Wong.** *Inquiry.* 2000.
  
- Tracking the state children's insurance program with hospital data: National baselines, state variations and some cautions.  
**Friedman, Jee & Steiner.** *Medical Care Research and Review.* 1999.
  
- Does managed care affect the supply and use of ICU services?  
**Friedman & Steiner.** *Inquiry.* 1999.



If you have questions about HCUP in particular, you can write the agency at [hcup@ahrq.gov](mailto:hcup@ahrq.gov). You can access HCUP products by going to [www.ahrq.gov/data/hcup](http://www.ahrq.gov/data/hcup). The direct link to HCUP Net is available at [www.ahrq.gov/data/hcup/hcupnet.htm](http://www.ahrq.gov/data/hcup/hcupnet.htm). The central distributor is available at 866-556-HCUP (4287) toll-free, or writing [hcup@s-3.com](mailto:hcup@s-3.com). Say you just want Arizona, Colorado and Utah, and you want the hospital data. You have several things you can do. You can call the states directly and ask them for data, but if you wanted any uniform format, you can go to the central distributor. The central distributor will process that request directly to the states and provide the data so you don't have to go to each state agency. It's one-stop shopping for state data.

Finally, I'd like to talk about the health Web site locator module that NAHDO did starting in 1998 with the SOA. NAHDO had demand from its members, as did the SOA, quick links to health Web sites that somebody had looked at. They didn't want a link to a search engine that gave them a whole list of key words, but rather one that directly linked them to health Web sites so they could search and catalog these sites. It's taken a while, and it's taken several versions, but Table 2 is the latest. Today, we're here to talk about what would be helpful to you.

Table 2



This is how the site looked a few weeks ago, and I think it still looks this way. There's a search engine, and I typed in Health Insurance Portability and Accountability Act of 1996 (HIPAA), and if any of you haven't heard about HIPAA, come see me afterwards. These are the different health topics you can link to. The HIPAA page just brought up in the access database a listing of the different HIPAA sites that are in the database. One of the challenges that we have is populating the database. Every day I learn about three or four cool sites.

We've made it so that others can add a site, and we have a template. So, if you see a site that you think it should link to, we've tried to make it user-friendly so you can multiple-code a site. The hardest part, and it took me a couple of years to figure it out, is that HIPAA may also be other things. It may be a database or it may be an article, and it's really tough to catalog those. It may even be a chronic disease, such as a cancer. You can multiple-code the sites from the remote entry as a chronic disease or a public health database and submit it.

If you need state data and don't want a whole state's in-patient data,—(sometimes you might just want a subset of a subset) NAHDO maintains a network of state data agencies across the country so we can put you in touch. Say you just don't want to buy a whole dataset, but you want a subset of that data. We can put you in touch with that state person who is responsible for the data. NAHDO is the National Electronic Disease Surveillance System (NEDSS) partner. With the advent of 9/11 and bioterrorism, there's quite a bit of money flowing in to develop public health surveillance systems and public health data systems. So all states are virtually involved in public health preparedness activities, not the least of which is the



NEDSS system development, which is an architecture and data to integrate public health databases at the state level and speed up reporting to the CDC. The reason I mention this is it will significantly change the way states capture and report data, and hopefully we can get timelier and more integrated data at the state and federal level because this is a major initiative. By 2002, public health and social services emergency funds are funding state and local public health capacities, including IT and surveillance.

NAHDO, in addition to tracking state and federal data trends and linking people with data across the country, is developing an emergency department tool kit. We just had a conference in Washington, D.C., to bring state and federal agencies together to talk about the next steps in emergency department data development. We are developing a resource kit and clearinghouse; we will eventually have the software tools and links for statewide emergency department data collection and use on our site. We're also working with five states and Stanford to build research and infrastructure capacity, and pharmacy data will be a part of that. We have an Internet query system that's also part of the Data Web and the health locator site for the actuary. I just thought I'd give you a little flavor of what NAHDO does. We also provide technical assistance to states as they develop their data. Many states, when they get a mandate to collect hospital data, don't get very much money. They say they have to do a statewide data development, and so NAHDO goes in and has legislation model rules, a model data submittal manual and even SAS programs that we pirate from other states. We're able to help that state develop its data system on the cheap, so to speak, so that it's up and running hopefully within a year or two after that mandate is in place. That's where I spend a lot of my time.

Again, I ask you for feedback and remote data entry. I hope this presentation has provided you with some resources that you can use in your work.

**MR. CATTERALL:** I'll begin my presentation on the Health Web site search module. How many of you have used or at least accessed the search module? I see from the show of hands that there are a fair number of you. How many of you completed the survey that was e-mailed to everyone? I don't see as many hands. Well, we need your input on this.

I'm going to be talking a little bit about the history behind the search module and talk about what's on the site. I'll give one example showing the practical, real-life application, and then I'll talk about the survey that I had hoped everybody had filled out, but at least some of you did. Then we'll talk about future possibilities for the search module.

The idea actually dates back all the way to 1988. Someone at the SOA, I'm not sure who, decided that we needed to have an Internet database locator, and, as Denise mentioned, many people in other organizations have felt the same way. I'm not sure what happened between then and 1996, but by 1996 NAHDO had received a grant from the Robert Wood Johnson Foundation to develop the National Health

Information Resource Center (NIHRC) website. In November, NAHDO received a grant from the SOA to expand NHIRC's content.

The NHIRC website provided links to more than 150 health data resources on the Web. It was really a terrific site. I used it a few times and was overwhelmed by the amount of data that was available. In August of 1997, NAHDO submitted a grant proposal to the SOA for enhancing the NHIRC site. The three enhancements that they proposed were a keyword search capability, a health database locator and links related to outcomes and performance measurement. The SOA approved the proposal, so NAHDO made the enhancements while also moving NHIRC's content to a new page on its Web site ([www.nahdo.org/soa/soa1.html](http://www.nahdo.org/soa/soa1.html)). If you haven't visited it, do take a look. As a result, the Health Website search module essentially replaced the NHIRC. The "nhirc.org" domain name was released and is now used by some other organization as a general search engine.

As I mentioned, I'm chairman of the SOA/NAHDO Work Group subcommittee of the JCCHI. The work group's charge is to review the search module on a regular basis and, as appropriate, provide feedback to NAHDO regarding desired changes and enhancements to the module that will allow for its continued usefulness to the SOA's health practice membership. One of the things that the staff at the SOA mentioned when this group was formed was that, because of the dynamic, ever-changing nature of the Web, it's necessary to have somebody looking at tools like this on a regular basis to make sure the module is up-to-date. Currently, the work group has eight members, but more are always welcome. Send me an e-mail if you're interested in joining.

Now I want to talk about the organization of the search module. There's the keyword search. There are links organized first by health topic and then by type of site. There are related links, several to NAHDO, and there's a link to the SOA Web page. There are links to other search engines like Google and Lycos, in case you can't find what you need on this page. There are also opportunities for user input, as Denise mentioned. You can add a site. You can report broken links. If there's an outdated link or a link that doesn't work, you can report it to NAHDO.

One of the health topics that was particularly important to the original grant proposal was the outcomes and performance measurement page. If you go to the bottom of this page, you'll see there's another page of information. Some of the categories have many, many pages of information and of links. Others just have one or two. If you look under type of site at the databases category, there are 15 pages of these links. So, there are a lot of databases that this tool links to.

Here's the application example I wanted to review with you. This was a task I faced a couple of months ago. We were working with a state Medicaid program, and we needed to estimate their future expenditures for hospital care. What we wanted to do was to go to the CMS Web site to see what their projections were for future Medicaid expenditures on hospital care.

I entered "Medicaid projections" into the keyword search. It brought up a link to the former HCFA site, which actually, if you click on that, takes you to the CMS site. There's a link on that site to "stats and data." Click on that, and it results in a list of the statistics that are available, including national health expenditure projections. When I clicked on that, I found the most recent projection going through 2011, published in March of 2002. I clicked on the tables. I went down to Table 6, Hospital Care Expenditures, Aggregate and Per Capita Amounts by Source of Funds, meaning out-of-pocket versus private health insurance versus various public programs. Here in the last column, in the lower right-hand corner, are the projected expenditures for Medicaid programs nationwide for hospital services. So, eventually I was able to get to the data I wanted. As you can see, it took quite a few clicks to get there. If we did what they call a "click stream analysis" on this, we wouldn't score too well.

One of the reasons we developed and sent out the user survey was because we felt there was room for improvement. We wanted to get feedback from a broader cross-section of the SOA Health Section membership than just the people who are in the work group. We're seeking answers to questions such as: What are the typical users particular needs (whether that person has used the search module or not)? How helpful is the search module in meeting these needs? How can the search module be improved?

In the e-mail that you got about a month ago, we asked whether you'd accessed the Health Web site search module before, and we invited input from those who hadn't accessed it before just to give us information about the tasks for which you were often in need of data. If your work involved health insurance products or health benefit plans, what specific area did you work in? In other words, what kind of health actuary are you? What type of data do you most often need? Claim costs? Incidence? Prevalence? Utilization? Unit cost statistics? We continued on down to "information on new drugs or clinical research", and there was a space to provide other types of data if you didn't see it in this list. We wanted to find out if you found the search module to be helpful in finding the data you needed. We wanted to find out if the categorization of Web site and if the keyword search were helpful.

Also, we were curious to find out whether or not you thought the search module was doing a better job at getting you to the data you needed quickly and easily than general search engines. We wanted to find out if you'd recommend the module to other health actuaries. And, finally, there was one question for which we hoped we'd get a lot of feedback: "What changes would you recommend making to the search module to make it more useful to health actuaries?" The responses to the survey have been received, but they haven't been tabulated yet, so I can't give you any information about what we found out from the survey.

Now, I'd like to talk about the outlook for the search module. First, we have to keep in mind that it's a work in progress. There are a lot of out-of-date, broken links. There are duplicate listings that need to be deleted. I think having descriptions for

the linked sites in our search module is very helpful, but some of them need to be revised or expanded. Some links should be moved to different categories to make them easier to find. As Denise mentioned, a lot of links have to be listed in several categories. There are also a lot of new links that need to be added. Every day you can find a new site that will be useful to health actuaries and others who work with health care data.

The keyword search needs to be improved. One of the work group members mentioned looking for information on organ transplants and the cost of that. If you enter transplant into the search engine, you get zero links. If you enter organs, you get many links. We want to make it so that you can enter any keyword that's related to the topic and get the whole list of links. Search results often omit links that should be included. Links are often not specific enough, such as the organ transplant example I just gave. When I entered my search terms it took me to the HCFA or CMS home page, but it didn't get me to the specific dataset that I wanted. We could revise the search engine so that it gets you directly to the dataset that you need so that you don't have to go through so many clicks.

We need to revise and expand the Health Topics list. I think we need to include a few more topics of interest to health actuaries. This is a tool that's used by many people who are not actuaries. Denise was telling the work group earlier that there are certain categories that need to be included in the list of topics by NAHDO's other member groups, but we certainly want to make sure that health actuaries can find the information they need. We need to include subcategories on the topic list where appropriate, such as under Health Insurance, including plan design, regulations, small group market, etc.

Something that would be very useful would be to have user ratings or reviews of the linked sites so that we could get ongoing feedback about which sites you find most useful. That could be used to prioritize the links so that instead of getting an unprioritized list of links that might go on for many pages, you'd get something that lists them in decreasing order of priority, either according to hit count or usefulness as rated by the members. It could also be prioritized according to the number of links from other sites. As I understand it, that's how Google works. I also think it would be helpful to distinguish between different types of sites, such as free sites versus pay sites, and sites where you have immediately accessible data versus sites where you encounter essentially advertisements for data that you can order. Sometimes you need to get the data immediately, but sometimes you can afford to wait for a few days for something to come in the mail.

I have a couple of final thoughts. Some of these changes might require additional funding from the SOA, so we'll need to set our priorities carefully. It hasn't really been worked out whether it'll be primarily the SOA or primarily NAHDO that will be maintaining the site, especially maintaining its usefulness to actuaries. That's another reason why I'd love to have more people in the work group, because it's a big project, and we need as much help as we can get. But even if you can't

volunteer your time for the work group, we do want to get your feedback. Feel free to let us know how you think the search module can be made most useful.

**MR. RICHARD TASH:** With all the public databases that you're collecting data for, what is done to audit to ensure quality and consistency of the data that's received, whether it's from all the states or wherever the source is?

**MS. LOVE:** I'll speak about what I know best, which is discharge data. That includes the 44 states or so that collect statewide discharge data. It would be rare for a state to go back and audit against medical records. Most states have systematic edits, so that systematic edits that find missing or clearly wrong values outside of the set parameters will be sent back to the hospital. In fact, that's what takes a state so long sometimes. The data is sent back to the hospital until the error rate is less than, in most states, two percent of systematic errors. There are also some coding checks, and some states use some of the 3M tools to check for gender conflicts, age conflicts and outside value conflicts. Most states are not doing a validation check with a medical record, because either they cannot do so legally or the funding isn't there. There are some states that have done so, such as California. CMS has a report out stating that they went back to the medical records and to the claim state and didn't find as many errors as you would think. So, administrative data has its limitations. I would love it if we could come up with a minimum audit protocol that we could apply as standard across all states, but we're not there yet.

**FROM THE FLOOR:** One of the few sites you mentioned that I've been to is the CMS site with national health expenditure projections. I looked at it last year with the data through March of 2001 and then again this year through March of 2002. One of the things that I noticed, and I was looking at the per capita expenditures by source, is that this is created and maintained by CMS, and, like a lot of these sites, they're governmental. I was curious if any of this data is "political" data, particularly the forecast, and might be subject to some political manipulation. One of the things I noticed with that projection was that the increases in per-capita spending forecasted for Medicare as a payer decreased for near-term years while the ultimate was still pretty high. We're all concerned about potential cost shift from Medicare in the future. We know it's occurred in the past, and it's occurring at the present, but it seems to me that that data seems a little rosy in terms of per-capita increases in Medicare spending going out the next 10 years. I wonder how valid or how subject to influence that data is from political sources.

**MR. CATTERALL:** Just from my perspective, CMS is not a particularly political organization. However, CMS and the Social Security Administration (SSA) do have to make projections that are subject to variability. They come up with different scenarios. The projection that I cited here was their intermediate scenario. There has been some discussion in the past about whether, for example, the SSA's intermediate scenario is, in fact, a mid-line estimate or a rosy scenario. I think it's important, though, to keep in mind that Medicare, unlike private payers, has

complete control over the prices it pays. It doesn't have control over utilization, but they can just say they're not going to pay providers more than X amount of dollars or more than X percent more than what they were paid the previous year, and, indeed, they can even mandate decreases in reimbursements in certain cases. That might be one reason why you see what looks like unusually or unexpectedly low expenditure projections. Whether there was any influence brought to bear I really don't know. That is a question for someone who actually works for CMS.