NATIONAL HEALTH DATABASES

Moderator:	RICHARD A. KIPP
Lecturer:	DENISE E. LOVE*

Society of Actuaries (SOA) research and activity on health databases will be presented.

MR. RICHARD A. KIPP: I'm a consulting actuary with Milliman and Robertson's Radnor office, which is just outside of Philadelphia, Pennsylvania, and we are here to talk about a couple of things. We're first going to spend a few minutes talking about the Society project that was conceived a little over a year ago and is having just a little bit of difficulty getting off the ground, but we hope it will do so over the course of the next couple of months. Once I update you on the status of that project and give you an idea of what the project is, we're going to move on to a discussion that actually should prove to be interesting. This will be given by Denise Love who is here with us from the State of Utah. I'll introduce her more formally later. She will talk to us about an association of health data organizations that has formed over the course of the last few years. Then she'll give us her view on the development of a database. She will discuss the issues and concerns that come into play as you're trying to put together one of these databases.

I'll give you a little bit of a background to kind of set the stage for why the Society was interested in doing this project in the first place. Any presentation I give about anything these days usually starts with some mention of cost. You know the cost issue is something that has pushed the entire industry in the direction of managed care. It has increased the number of players that we're finding in the health care business these days. It has also increased their interest in data and doing analysis of the quality and cost of the use of health care services.

Without question, quality is an important issue these days. It's something that people are paying a great deal of attention to, and we're trying to come up with reasonable methods of measurement. Data are at the heart and soul of this issue. Quantification and measurement are extremely important with regard to quality and cost.

The environment that we're working in is very competitive with numerous players. All are looking at price and quality as the dominant issues. They are competing in a completely different way than they have before. We're looking at many niche players that are coming into being. There's a new kid on the block—the pharmaceutical companies that are becoming more and more interested in playing a part in actually managing care via disease management. I'll talk a little bit at the end about disease management and its place in this new environment.

The medical infrastructure and data necessary to identify which providers are providing the right and appropriate types of service to the various patients that are entering the medical system is of great concern. A number of companies have begun the process of establishing themselves as data processors and information managers. Due to consolidations, new

---

*Ms. Love, not a member of the sponsoring organizations, is Director of the Office of Health Data Analysis at the Utah Department of Health in Salt Lake City, UT.

companies are being formed. Data are their primary interest and data give the ability to providers to analyze themselves and see how they compare to standards and so forth. It's becoming a very data-intensive, information-intensive health care society.

The physicians, the providers in general, are organizing themselves into entities called integrated delivery systems. You've probably heard others talk about them. You know the alphabet soup. These organizations are forming so that providers can take on risk and, form groups that have a little bit more power in the world of delivering care and financing care.

The things that organizations of providers need to do well in this world are listed here, such as medical management. If you're a provider group you have to have some sense of how you are meeting standards and what the best practices are. Designing medical protocols is something that many companies are attempting to do. To perform this function for the different entities that are in the health care system, one needs information. This is why having access to data and creating the Society's database is so important for us as users of that data. The Society's idea of how they're going to create data access is not exactly what some of the states are doing, but they are data, nonetheless. It will provide us with a great deal more information for the future to help organizations, providers and others make decisions about risk. There's the outcomes, management information systems, financial management of capitation programs and all the things that you're well aware are going on in the marketplace these days in the health care field.

The business of disease management is something, as I mentioned, the pharmaceutical industry is particularly interested in. Talk about a data intensive kind of an issue, I mean this is adding another dimension to what we've typically looked at in terms of cost and utilization by component and requires organizing those same data by disease. You can imagine the difficulty involved in defining a disease. There's the scope of service issue. There's a definitional issue that is fundamental to doing business in this fashion, which is extremely difficult and can't really be done well unless you have a lot of detailed data available to conduct studies and analysis.

The pharmaceutical industry's point of view is a little bit different than some of the other industries that have played a part in the health care systems. Certainly they are a part of the system today, but are really only suppliers. They're commodity producers and find them-selves well at the end of the spectrum of control of the management of the care that's being provided. They're wanting to move themselves forward to integrate themselves. To do so, they need data and access to patients. They have to have information that gives them the ability to form opinions about the potential cost and use of services for the types of diseases for which they have drug therapies. This is needed so they can make educated guesses as to how to contract with the various managed care organizations that may use them for assistance. So there are many interested parties all looking for very detailed data.

Now the Society is about to undertake a project. We'll look at its purpose and its research objectives. The purpose as the Society sees it in doing this project is to provide reliable health care data to the users of that data, primarily the actuaries that are members of the SOA. They'd like it to span major markets both public and private. They'd like to establish the actuarial profession as the primary source of this kind of information. This is an issue that haunted us as a profession as we dug deeper and deeper into the health care reform problems of a couple of years ago. Actuaries weren't always thought of first. We weren't always consulted first about health care system issues, and we want to have that changed. So that's

one of the main reasons that the Society is interested in putting together this source of information. They'd also like to have it reduce data source research time which is, obviously, a waste of time in the industry. Every one of you who wants to go out and find out some information about a particular issue ends up doing it on your own and kind of recreating the wheel every time you do that sort of primary research. It's unfortunate, but hopefully, through this particular project, that can be avoided to some extent in the future.

The research objective is to have the data set include cost and use data, health care utilization, cost of claims, the price of services, outcomes (to the extent possible), incidence, survival rates, some information about the kinds of coverages that are available, trend information and other related types of things. Obviously, we're also going to need some exposure information. That is the population base that relates to the claims cost and utilization data. We'll have to have population statistics and some demographics and potentially other characteristics of the people exposed to risk.

The SOA would also like these data to be very broad in scope to assure an extensive resource. Also there's an issue here of when and how to update the data set. Making it a dynamic, living, breathing data set is really what the intent is.

The request for proposal is going to include these elements: (1) What data are to be in-cluded—it's obviously important for people that are going to help the Society in doing this project, (2) who the intended users are, (3) what form the data are expected to come to the Society in, (4) what, if any update process there will be, (5) what disclaimers may be neces-sary, (6) what potential enhancements are needed—that is, things that we might like to see happen in the future but aren't necessarily part of the project at this point, and (7) the proposal contents themselves.

The data would include almost anything that you can imagine like studies and publications of the U.S. and state government organizations. The National Center for Health Statistics (NCHS) can provide a wide variety of data sources, surveys and information from various things that they do on a regular basis. We would want to identify all these things and make it clear to the people that would be using this reference system exactly what these data sets are. Databases that might be available from professional associations such as the American Hospital Association (AHA) and the American Medical Association (AMA) would all be included. Information published in some of the journals, (those juried and nonjuried would include things done by professionals and other interested parties in the health care world) would be included. The other sources, obviously, would be books and publications from foundations and the private research groups. There's a lot of research that's being done these days that addresses many of the issues, especially outcomes.

We see many studies being done by universities and other interested parties in the outcomes area, and this is, again, something we want to catalog and have available for people to tap into. Obviously, actuarial publications, material published by consulting firms, insurers, and service providers could be included. Any kind of study that hits on a health care issue is a potential member of this data set that the Society's putting together.

The users, are primarily people in the actuarial profession, but are certainly not exclusively those. There's no reason to think that others couldn't tap into this resource and make use of all the cataloging that will be done, but I think the Society definitely sees it as a primary resource for the actuarial profession.

The form of the data that the Society is going to be looking for will be an abstract, a catalog or whatever you want to call it, but it's not the primary data source. The Society doesn't see itself as a primary data warehouse (at least not yet) of very detailed, machine readable kind of data. You know they just don't want to do that quite yet. It could be very expensive to collect that information and maintain it and update it. At some point in the future, it may be feasible, but at least right now the idea for this particular data set is to have it be an abstract system; that is to have some sort of identification of what the source of the material happens to be, the title of the publication, the date published, a brief description, etc. If there's a record layout or anything like that information, they would like to have that catalogued and be readily available to anybody that's interested.

The way of initially communicating this information would be to do it through a bulletin board type of process. The Society has already taken steps to try to establish that kind of process with other systems that they have in place. They'd like to add this catalog device to that sort of a system to make it easy to access. Then somebody could just call in and tap into the thing at any point in time.

The update process would need to include the maintenance of the old sources of data but, also, new information that becomes available. There are going to be many new studies that are done—things that are not necessarily updates of old bits of information that may have been collected through this initial process. The Society would like to have the best available information in both regards at all times.

The Society is also not interested in necessarily taking ownership for any of these data. Accuracy and appropriateness of the use of any particular data set is really up to the user of data. The user needs to check the data out. The user has to work with the data to discover its limitations. There may be some notations in the SOA data catalog about potential problems and appropriate uses of the information. The Society can't take responsibility for every data set that it's going to have in this warehouse. So users are going to be responsible. Also the nature and scope of the data sources can actually change over time. Any of you that worked for insurance companies have seen this with data sets that you have created for yourselves. The claims data sets in particular are subject to all sorts of problems. For instance, a system change at a particular point in time can suddenly cause variables that you were very comfortable with to take on new meanings. You may have in the past had claims processors that were filling in bits of information on a claim form. These data may have previously been key-punched and now, in this new automated world, you may have machine readable data transferred through a tape-to-tape process. A change such as this can change the quality and quantity of the information that's available very quickly. The data users will need to be aware of those kinds of issues. The Society will try to track these things to the extent that it can, but the user must beware.

Out of the enhancements that the Society sees happening in the future (and these are things that are not necessarily part of the initial proposal), is the publication of some sort of a data source book, so that all of the information that's out on this bulletin board would then be put down in hard copy and potentially sold. I'm not sure exactly what the method of distribution of this book would be. But the other more interesting aspect of this would be to eventually have some of the detailed data available through some on-line process. In this scenario you could just call in and actually have detailed data sets, like an inpatient data set from one of the state inpatient discharge systems, available for your use. It's available so that you could tap right into and copy it and make use of it and do whatever analysis that you're interested in

doing instead of having to first find it through the catalog, obtain it and then start working with it. It would shorten the process considerably. But, again, this is not something the Society sees happening very soon and certainly not right off the bat.

The proposal will probably be distributed to potential respondents fairly soon. There is a draft of the request for proposal (RFP) that has been submitted to the committee members. The RFP outlines an approach which identifies issues and some of the techniques that would be used by the respondents. Any limitations that respondents see that need to be identified would, obviously, have to be part of their response. Clearly, a cost estimate is going to be fairly important to the Society, because it doesn't have unlimited funds to use for this or any other project for that matter. The Society would like to see the things broken down into component parts, computer salaries, manuscript preparation, etc. The Society wants to see the steps and the costs that are associated with those steps in detail. Also, the Society would like a proposal schedule. They want the respondent's ideas on the form and on the distribution material. Resumes of any of the researchers, project teams, and so forth will be required. These are all things that would be expected and if there's evidence of other capabilities that the respondent may have that don't happen to fit these particular categories, the respondent is welcome to submit those as well.

Our goals for this particular project are to issue the RFP within a month or so. I have the draft of this proposal. It has been reviewed at least once, but I think the committee needs to go back through it one more time. The RFP looked like it was in good shape, so it could come out within a month or two at the latest. It can then be put out on the street so that we can get the responses back to allow the evaluation of the bidders and the implementation of the health care data source catalog.

It has taken some time for the Society to rise to this, but I think that it's definitely on the right track. I think the profession needs this sort of tool. I think it would be a help to many of us that do this kind of research on a fairly regular basis to have all this information in one place. Hopefully, the Society will be able to get this project going and implement it this year.

I would like Denise Love to talk to you about the National Association of Health Data Organizations (NAHDO) as well as the State of Utah's data set. Denise is an R.N. and an MBA. She is the Director of the Office of Health Data Analysis at the Utah Department of Health and she's an appointed Executive Secretary of the Utah Data Committee. Denise has done a lot of work with the data in Utah, and when I talked to her before the meeting, it sounded like she has some interesting ideas with respect to what the state is going to be doing with data.

MS. DENISE E. LOVE: As vice chair of the NAHDO, I will speak on behalf of other state, federal and private data agencies, and I'll share a little about what Utah is doing with statewide data.

As I listened to Richard Kipp's presentation, it occurred to me that the information available is growing almost exponentially, and it's tough to keep track of it all. So instead of health data analysis, it's almost a meta analysis when many types of data are used for research and decision making. With so many sources of data, it makes sense to pool our resources.

The insurance and provider industries recognize the importance of data. Policymakers at the state levels are beginning to recognize the data gaps and may not realize how difficult it is to

get the needed data. Investment of limited state funds in data system development varies between states, and the long-term nature of establishing databases will exceed the terms of many policymakers. So this is a real challenge.

STATE DATABASES
I pulled together a summary of federal and state databases, based on listings from several sources. Many states have public data sets or reports with patient demographic, diagnostic treatment data, provider characteristics, claims data, and health practice measures, in addition to vital statistics data. Prices of public data sets for hospital discharge data range from a low of $100 to approximately $20,000. Access to these public tapes varies by state.

Major challenges to statewide data development are data quality, privacy and confidentiality, and comparability and release issues. Laws and policies in collection requirements and disclosure varies between every state. Additionally, states struggle with the balance between timeliness versus accuracy. Many states release data more than one year, and often two years, after date of discharge. Comparability and data quality are important, as combining data from multiple data sources is difficult and coding practices do vary. Due to privacy issues, the one data element that varies the most between states is the presence or absence of a unique patient identifier. Some states restrict provider identification as well.

There are 39 states with a mandate to collect health data. As states and business coalitions formed to collect comparable data, states were "doing their own thing" and uniformity and access policies varied widely. Ten years ago, the NAHDO was established by the Washington Business Group on Health and the State Intergovernmental Project to provide a clearinghouse of information and offer some guidance for the uniform development of core data elements.

NAHDO's mission is to improve health care through the collection, dissemination, public availability, and use of health data. The membership is quite diverse, and now includes Milliman and Robertson as a new member. As a nonprofit membership organization, it also includes many state and federal data agencies.

NAHDO provides technical assistance to states just starting development of a health data reporting system and guidance as well. For instance, many state legislators want to see immediate returns when state funds are invested, but data systems take many years to develop. NAHDO provides a particular perspective to affirm that this is a long-term project and is able to provide information on how other states have approached similar challenges and reminding all to engineer in time. Data integration is becoming more of an issue with states as well, and NAHDO can provide referrals to members dealing with similar issues.

Private sector involvement is critical and use of the data by the private sector is key to survival. However, interpretative reports can cause difficulties. Relevance in reporting is challenged due to funding structures. If a state is dependent on legislative funding, it has to be careful not to get the provider community too upset. So use of the data in credible and meaningful ways is always a challenge for the states. Through NAHDO, sharing of "best practice" in reporting is also a key feature.

NAHDO supports the development of a nationwide, publicly controlled infrastructure with partnerships. The public sector can provide a means by which proprietary interests are mediated and protected and the private sector is both a generator and user of data. With electronic data interchange and American National Standards Institute (ANSI) standards, the

possibility of combining administrative data with clinically relevant data may not be too far away. Currently, the cost is prohibitive on a statewide basis.

HEALTH CARE UTILIZATION PROJECT 3 (HCUP-3)
One example of a federally funded partnership is the Health Care Utilization Project 3. The Agency for Health Care Policy and Research (AHCPR) is making this national database available this summer. NAHDO served as a subcontractor to bring together the states and worked with Systemetrics as a technical subcontractor as well. The objectives of HCUP-3 are to provide uniform inpatient data with linkage potential and capacity for research.

There are two types of databases now available in HCUP-3: National Inpatient Sample (NIS) and state inpatient database (SID). The SID includes 100% of hospitals and discharges in 12 states and is available through each individual state and subject to release policies of each state. The NIS is a sampling of 20% of hospitals in 11 states (excluding New York). The data are released by AHCPR.

The time period is 1988 through 1992. The number of hospitals included in the SID are 2,000 with a total of 72 million discharges and the NIS contains approximately 850 hospitals with 42 million discharges. There are three levels of data elements. There are core data elements, that are fairly common between states, supplemental data elements, and developmental elements that vary the most between states. Examples of supplemental data elements include data quality indicators and physician identifiers. Developmental data elements are the most confidential.

Tape formats between states should be standardized—but may still vary between states in the way they are collected or compiled. For instance, payer fields, E-codes, and physician specialty coding vary greatly in how states require these data to be submitted.

The NIS is a stratified sample of hospitals which is longitudinal. Twenty percent of U.S. community hospitals and 100% of discharges from those hospitals are included. It will be updated annually and additional states may be added. Personal identifiers are removed and hospital identities are protected. CD-ROM versions will be available and the cost of an NIS tape is $800 per year of data. Contact the National Technology Information Service and they will send it after the requestor signs a data-use agreement.

As you are aware, some of the uses of an inpatient discharge database are to measure the use and cost of hospital services, to study medical treatment variations, to determine the quality, in some cases, of providers. Administrative data are limited in measuring outcomes, but are being used by states as morbidity surveillance systems. In Utah, we are putting hospital discharge data out on the Internet to support interactive queries, graphics, and tables.

One of the major limitations of hospital discharge data is that it just measures end use of medical care utilization. Factors such as patient health status, medical practice patterns, economic incentives, and training, are often not taken into account when variance is observed. States realize that there is this data gap, and states are expanding to outpatient data sets, adopting severity adjustment tools, and linking morbidity and mortality databases. Performance measurement for system accountability statewide is just developing.

With managed care, there is an opportunity for the public sector and the private sector to work together to improve the health status of enrolled populations. Risk adjustment is one area in

which actuaries and insurers may be well ahead of the public sector and population-based assessments of health status is one area in which public health specializes—it is an opportunity for sharing expertise and knowledge.