Health Section News

April 2000 – No. 37

In Search of That Most Elusive Companion to the Health Actuary . . . Data

by Richard A. Kipp

Anyone that has had the pleasure of working as a health actuary for any organization in the health care delivery or insurance industry has been faced with one or more of the following questions:

1. What will the cost be to provide a new benefit?
2. What will the future use of a particular service be?
3. How much could we reduce our costs if we change....
4. Are our members, patients, or policyholders sicker than average?

The list of course goes on, but these questions explore three primary drivers of health care cost change — cost, use and mix of services provided to patients.

To answer these questions, the actuary needs the inquiring mind of a tabloid reader and a data set that is ready for mining.

To be sure there is an enormous amount of data available to actuaries outside of what may be available from the operations of the organization for which they work.

Internal Data

Internal data sources are essential for the actuary to use in forming any opinions regarding the questions outlined above. However, there are many times when the data supply can not be used as the sole source of an analysis. Examples of such situations might be entry into a new line of business (e.g., Medicaid), an insufficient amount of data (e.g., first year of a new product's life), or benchmarking the organization against industry or market standards.

The most difficult of these situations is when an organization's own data is simply too sparse to allow one to use it without testing it for statistical credibility. Considering one's own data a statistical sample of a larger population of data is not a natural way to think for many people (especially the senior staff at your organization). To develop confidence in your analysis of such data, one should calculate some vital statistics, such as the mean and variance of the quality of interest. Once this is done, you can begin to understand how large a data set you would need to obtain a statistically sound estimate of the measure in question.

External Data

Once one decides that external data is necessary, the question becomes where to look. As most readers will have already discovered, no single source contains all the health data that an analyst may desire. The health system that produces the data is fragmented by geography and the self-interests of its participants. Having said that, there are two forces at work that at times seem to work counter to one another in moving us to a world with more detailed data access. The first is an overwhelming desire by states to assure quality care for all. The second is the privacy of that data.

The desire for advancing public health (along with other factors) has moved many states to require hospitals to report all of their inpatient discharges for all payers to a central state repository. The states then "clean" the data and use it for public health and health systems accountability purposes. Many of the states also package it for sale to interested buyers. The variables contained in these files are very detailed. Typically the file is a subset of the uniform billing (UB92) claim information that would be used to bill an insurance company or the Health Care Financing Administration (HCFA). State discharge data sets usually include patient demographics, patient status (discharge status, admit source), utilization and charge data, and diagnosis and treatment information. About 28 states collect physician identifiers, but not all of these states make this publicly available. In addition, most states include non-billing state fields with the discharge data, such as birthweight or gestational age of newborns, present on admission indicators, severity scores, and readmission indicators.

States generally prepare a record-level "public use file" that consists of de-identified or non-confidential data and a research file for public health and approved research. The public use files protect patient identity through encryption and aggregation of sensitive fields (e.g., date of birth is mapped to age grouping, date of admission to quarter of admission plus length of stay). Currently, 44 states collect inpatient discharge data. Twenty-six states are also collecting ambulatory surgical data; 13 collect ambulatory surgery data from free standing facilities. Eighteen collect emergency room data and 11 collect non-surgery, non-emergency data from hospital outpatient departments. Not all of the states allow all potential users access to all data, but the trend is in that direction. As mentioned, due to concerns about confidentiality, you cannot uniquely identify patients in the state inpatient discharge files. Therefore, analysts are limited in their ability to track people across time. However, in some cases you can link re-admission to original admission and by doing so, begin to look at crude treatment outcomes, such as re-admission for infection after surgery. One can also use a field that records discharge status to make observations about the use of sub-acute care in the treatment of various illnesses and also what the mortality rates for various conditions happen to be at various facilities. The state data is being enhanced and repackaged by some organizations. One example is the HCUP, directed and supported by the Agency for Healthcare Research and Quality (AHRQ). AHRQ develops and maintains inpatient data acquired from state government data organizations and private data organizations, which are primarily state hospital associations, to create a national health data information resource for research and comparative analysis. HCUP State Inpatient Database (SID) contains all hospitals and all discharges from 22 participating states. The HCUP Nationwide Inpatient Sample (NIS) database contains a sample of hospitals selected from the SID. It includes all discharges (about 6.5 million per year) from sampled hospitals (about 900 each year). The NIS contains fewer variables than the SID. For example, the NIS does not include detailed payer codes, which vary substantially across the states. Although the sampling frame is limited to participating states, the sample of hospitals in the NIS is selected to reflect characteristics of community hospitals nationwide. The NIS comes with weights that can be used to produce national estimates, regional estimates, and state estimates for participating states.

While 1998 data is available from many states, the most current HCUP NIS and SID data is for 1996. Even though HCUP, NIS and SID data is not as current as the state data sets, it is consistent in that all states' data used for the HCUP data set is mapped into a common format.

Another possible source is the Medicare data that HCFA makes available. Files referred to as BMAD and MEDPAR contain physician and Part "A" claims data for all Medicare beneficiaries. These files do not contain unique individual beneficiary identifiers, but do allow for analysis of many issues affecting non-HMO Medicare beneficiary costs.

HCFA also makes a data set available that includes the detailed claims data for a 5% sample of Medicare beneficiaries. This data set contains an encrypted patient identifier that allows the analyst to follow a patient's progress across time. As beneficiaries leave the Medicare rolls due to death or switching to Medicare Risk contracts, they are replaced.

The price for state inpatient data ranges from a couple of hundred dollars to approximately $30,000 depending on which state is involved. Missouri, Illinois and Pennsylvania tend to be among the more expensive. Vermont is one of the least expensive. Volume and thoroughness of cleaning tend to be the diferentiating factors. The HCUP NIS costs $160 per year. The cost of the HCFA data mentioned tends to be in the $1,500 – $10,000 range. The exact cost depends on the data sets requested.

HCFA also makes the Resource Based Relative Value Scale fee structure available. This can usually be downloaded from their site (*www.hcfa.gov*) or found published in the *Federal Register* each year after the update. A wealth of information, including the average adjusted per capita cost values used to pay the Medicare Risk Contractors, is available on HCFA's Web page.

New Data Search Project

In addition to the data available from government sources discussed above, are a host of private vendors of data provide information on both cost and utilization of health care services. To help actuaries and other analysts locate appropriate data from this multitude of sources, the Society of Actuaries (SOA) is working with the National Association of Health Data Organizations (NAHDO) to provide a Web-based resource called the National Health Information Resource Center (NHIRC).

The NHIRC was established by NAHDO in 1996 with grants from the Robert Wood Johnson Foundation and the SOA. The NHIRC provides abstracts of available data and links to a variety of Web sites of interest to health researchers, policy analysts, and actuaries. In 1998, the Health Section Council awarded NAHDO a grant to construct a search engine that would make it easier and quicker to find appropriate data. The SOA Committee on Health Benefit Systems Research (CHBSR) tested this search engine and made it available in on the NAHDO Web site early this spring. More than 500 Web sites are now included in this search engine, including all of the government sources of data discussed previously. The engine also includes Web sites maintained by some healthcare associations. These association sites typically do not contain as much hard data as an actuary would want or need. The CHBSR has instructed NAHDO to make it obvious to the user of the engine which sites contain hard statistics and which do not. NAHDO is also adding search categories for cost and use data as well as links to some additional web sites.

To access the NHIRC search engine from the SOA home page (*www.soa.org*), click on "Links to Other Sites" under the blue "FIND" heading, then click on "Other Pages of Interest." Scroll down and click on the "National Association of Health Data Organizations," then click on "Health Web site Search Module."

In an upcoming issue of this newsletter, we will explore some of the other sources and uses of health data, as well as the privacy regulations that are soon to be enacted. Suggestions for these discussions may be sent to Tom Edwalds, senior research actuary at the Society of Actuaries, (*tedwalds@soa.org*).

(Edited by Thomas P. Edwalds, FSA, MAAA).

Richard A. Kipp, MAAA, FCA, is a consulting actuary at Milliman & Robertson, Inc. in Radnor, PA. He can be reached at richard.kipp@milliman.com.