Data Mining

Stronger computer tools allow deeper analysis of medical research, patient care and insurance data.

by Mark Hagland

April 2004 - Healthcare Informatics

As director of the Pediatric Brain Tumor Research Program at Children's Memorial Hospital in Chicago, Eric Bremer, Ph.D., knows well the tremendous potential of data mining--using software programs for pattern recognition and predictive analysis. He's steering a project that demonstrates the future of data mining in medical research.

A few years ago, Bremer and his colleagues began to perform gene expression analysis for pediatric cancers. What makes a liver cell different from a kidney cell, for example, is its expression of genes, since not all of the 30,000 to 40,000 genes in the human genome are expressed in every cell, he explains. "So we're looking at patterns--I call them fingerprints--of a tumor, based on the genes expressed in that tumor. If we can accurately predict what's going on with a tumor, it should allow us to build more precise treatments."

With software from Chicago-based SPSS Inc., Bremer and his colleagues have isolated pediatric leukemia CD markers (antibodies that bind to proteins on the surfaces of white blood cells and leukemic cells) and hope to use the data to improve diagnosis and treatment of the disease. Chemotherapy and radiation doses could be far more precise and less physically devastating if physicians had more detailed information--particularly for brain tumors, which until now have been resistant to such precise diagnostic information management, Bremer says.

None of these recent advances at Children's would have been possible without data mining. Computerized programs can analyze data at a rate and level of complexity that is impossible to do manually. While organizations such as Children's are working on core clinical research, hospitals are also using data mining tools to analyze patient care data and case management.
Health plans are digging into data for better population segmentation and health management.

**New needs, new uses**

A gradual but fundamental shift is taking place, says Fran Turisco, a consultant in the Boston office of First Consulting Group. "Organizations used to mine claims information and financial data," she notes. "Now, they need to mine clinical information." Hospitals have lab and medication information but sometimes don't have a lot of electronic information with clinical documentation. "So they're putting in some of the advanced clinical information systems needed for data mining."

Hospitals are just beginning to realize the importance of data mining in all the key initiatives they want to pursue-computerized physician order entry (CPOE), patient safety, and prescription drug error reduction. "You won't see data warehousing, for example, at the top of priority lists on industry surveys" among hospitals, Turisco says. "You will see things like CPOE and patient safety." But hospitals need data warehousing to support patient safety and CPOE. "You need the ability to do ad hoc querying across a large enough group of patients," she explains.

Health plans that already use data mining have reasons to employ better data mining. Deeper analysis of member populations and better assessment of at-risk populations can reduce risk and better serve their members' needs, particularly as consumer-driven health plans become a force, experts say. Soaring medical costs are driving such improvements, says Charles Berger, senior director of product management for life sciences and data mining at Boston-based Oracle Corp.

Berger describes a typical first-generation use of data mining in which a payer client used Oracle's data mining software to uncover billing anomalies. "Certain patients repeatedly would go from doctor to doctor for procedures," he says, and "certain doctors [were] routinely running up the biggest bills with those procedures." Now, he notes, health plans are beginning to make more sophisticated use of data mining tools to do more refined segmentation of their member populations.

Some in the industry even believe the term "data mining" is outdated for describing the tools that are emerging to perform the newer functions. Suggested replacements include "predictive modeling" and "predictive analytics."

Cathy DeSesa, systems engineer at SPSS, says the distinction is significant. "Data mining is a very general term, but predictive analytics is really what we're doing. We've got a ton of data points, and we've been generating lots of data. What everyone wants to do is predict when a bad thing might happen and intervene.
before that." In predictive analytics, she adds, it's important to do "what it takes to get that good prediction, that piece of actionable knowledge, regardless of the product or the domain." That's true across sectors, from hospitals to health plans, she notes.

Waltham, Mass.-based IHCIS also offers predictive modeling, described as "a tool designed to predict both financial and clinical outcomes for every individual in a population for the next 12 to 24 months," explains Dogu Celebi, M.D., vice president of clinical affairs and client services. The goal is to identify appropriate targets for case and care management, he says, adding that with the advent of consumer-directed health plans, IHCIS is preparing a consumer tool for self-health-risk assessment as well.

Celebi says that health plans are getting ready to embrace data mining for real analysis, to "really push this information to the providers, and push this closer to their information process. A lot of plans are starting with small steps, funding projects to help doctors minimize prescription mistakes, etc. Everybody's recognizing the importance of getting to that point."

Research possibilities abound

Berger says the next frontier is medical research. Oracle is working on complex data mining in genomics, including projects at the National Institutes of Health, Bethesda, Md. With data mining, he notes, "We can handle hundreds, even thousands, of variables. So we can class by customers [and] can categorize."

The potential for patient care and public health has encouraged a few pioneering organizations to develop their own software programs. Medical researchers at Children's Hospital, Boston, developed an analytical system called AEGIS (Automated Epidemiologic Geotemporal Integrated Surveillance). It gathers and analyzes data from various emergency departments, looking for patterns, explains Kenneth D. Mandl, M.D., a Children's physician and an assistant professor at Harvard Medical School.

To date, nine Massachusetts hospitals (several of them pediatric) are feeding data into Children's database. Mandl and his team use AEGIS to analyze public health concerns, such as potential infectious disease outbreaks, unexplained rapid increases in pediatric asthma, and even lead poisoning in children.

The system, which went live in early 2001, has already tracked "naturally occurring phenomena, such as influenza, seasonal variations in encephalitis and meningitis, seasonal variations in gastroenteritis," Mandl reports. But the true innovation of the system is the speed of data availability. Patients coming into one of the participating hospitals' emergency departments are immediately assigned to a syndromic group. "We're able
to monitor activity across respiratory infection, gastrointestinal infection, hemorrhagic rash, neurologic syndrome," he says. "Then they are geo-located and plotted on a map.

The team also has developed algorithms for detecting abnormal trends over time and "abnormal clusterings of patients" in a region. The post-9/11 implications are strong, and Mandl says his team's work is tied into Massachusetts' overall biosurveillance initiative, which includes 50 state programs.

Challenges and potential

Hard work is ahead for data mining, particularly in the core programming and development needed to leverage its potential. Until recently, data quality and supporting technology have been the biggest barriers. Celebi says that "data quality has improved significantly, and technology has made giant leaps, though data availability and access are still giant areas."

Another challenge, Oracle's Berger notes, has been the Health Insurance Portability and Accountability Act's privacy mandates, which prohibit unauthorized use of patient-identifiable information. DeSesa adds that "a lot of mis-sing data," as well as disparate data sources, remain major problems for both plans and providers.

Still, the potential is great for health plans and providers to use data mining tools to improve patient care and plan member service. Public health, medical research, patient care, health plan member segmentation, population health management and individual customer services can all benefit. The question is, how quickly will healthcare leaders commit the interest and the resources that are necessary to fulfill that potential? On that, the jury is still out.

Mark Hagland is a contributing writer based in Chicago.