Accepted for Publication
Healthcare Data Management
October 2005

The New Frontier of Data Mining

By Kevin M. Fickenscher, M.D.

Healthcare is on the verge of a notable new period of discovery. With the advent of the electronic health record, new opportunities for uncovering patterns of care we did not know existed will come to the forefront of medical knowledge. As the healthcare industry continues its drive to enhance quality of care, promote services and reduce cost, these undiscovered patterns of care will become increasingly transparent, first for physicians, nurses and other clinicians, and ultimately for all consumers of healthcare.

The quality-service-cost triad has driven other industries to adopt new methods for deriving "business intelligence" from massive stores of available data. However, the healthcare industry is only now beginning to apply these proven intelligence-gathering technologies. In the electronic world of healthcare, we will be creating an abundance of data. Providers are now beginning to recognize the value of data mining as a tool to analyze patient care and clinical outcomes. Notwithstanding an increase in the use of information, healthcare is still far behind other industries in creating integrated, longitudinal data bases which can serve as repositories for data mining.

The Wal-Mart Method

For many businesses, the early pattern recognition and accurate predictive analysis that effective data mining delivers can mean the difference between profit and loss. Take Wal-Mart, for instance, and its accurate prediction about the demand for umbrellas in rain drenched portions of California this past winter. Through the analysis of historical sales data from stores in the region, national weather service predictions and the study of satellite weather patterns, the retailer accurately predicted not only the needed umbrella inventory, but the style and color most desired by its customers.

If we were to further explore Wal-Mart's capabilities in data mining, we would find that their system also is able to determine the best location for umbrellas in a specific store for the purposes of enhancing sales. Conversely, healthcare providers can barely predict which patients are likely to return to the hospital because of the patterns in their congestive heart failure.

That Was Then, This is Now

Over the last several decades, the healthcare industry has relied on secondary or extracted data for analytic purposes derived from two main types of data: claims information and financial data. The end result was that either the data analytics were confined to a data subset (e.g. pharmacy) or in an effort to create a more system-wide view, from secondary sources such as diagnostic codes, financial or operational data. Both approaches are widely used today to determine the impact of care delivery approaches, treatment modalities or service capabilities.

Such an approach was acceptable in an era constrained by the limited availability of clinical data and non-integrated information systems, but as providers deploy advanced clinical data systems, more granular, primary data is becoming available for analysis. Rather than accepting the physician's interpretation of data through the use of diagnostic codes or extrapolated financial codes, providers

will, for the first time, be able to objectively analyze the actual impact of one variable on another. The end result is that the healthcare industry is poised to discover new approaches in care delivery that consider a multitude of data points.

Looking Behind the Scenes

There are several drivers pushing the need for enhanced analytic capability in the healthcare industry. First, the **quality** movement is gaining momentum on multiple fronts and will not disappear. Demands from consumers, regulators, payers and internal constituencies are all requiring an ever increasing understanding of care delivery patterns in support of making changes to enhance quality, outcomes and patient safety. Increasingly, providers are warming up to the fact that data mining supports many of the key initiatives related to quality—computerized physician order entry, adverse drug event reduction and patient safety are only a few examples.

Second, the need for **value** in healthcare is painfully obvious. The demographic demands of our society, continuing double-digit inflationary costs of care, duplication of services and a host of other challenges currently outstrip the ability of society of pay for healthcare. While significant results have been derived from data analysis of the supply chain and from enterprise resource planning, it is the clinical front that offers unparalleled opportunity for creating value for the industry.

Imagine, for a moment, if providers held the ability to perform ad hoc queries using countless variables (e.g. past medical history, medications, treatment plan, protocol plan, social evaluation, housing situation, temperature, humidity and air quality) to predict which patients are likely to return to the hospital because of the patterns in their congestive heart failure. And, for health plans that are already beginning to use data mining to serve and monitor at-risk populations based on limited data sets, the need for value should provide even greater incentive for deeper, more complex data mining once clinical information becomes available. The potential savings that will be derived from faster, quicker, better care will far exceed the savings from all other sources.

Third, continued **consolidation** of the industry will drive the need for enhanced analytic capabilities. Large, geographically dispersed health systems will be unable to drive quality, support service standards or reduce costs without adequate knowledge of how care is delivered across multiple organizational fronts. These large systems are now attempting to collate their resources and determine the best strategy for creating standards of care across the entire enterprise.

Fourth, the availability of **data mining** software applications that can be used to uncover predictive patterns and relationships hidden in the data warehouse repository are now a reality. The data warehouse repository gives healthcare analysts a sophisticated analytic capability by providing the integrated, detailed data that has been extracted, transformed, cleansed, and modeled to support queries on pertinent business and clinical questions.

In addition, data dictionaries (also known as metadata) can be used to navigate thru the data, determine the definition of the data, and be leveraged in creating useful reports and analytics that respond to the needs of a wide variety of data users. Not only are these capabilities now both technically and financially feasible, but in a world that demands quality, patient safety, efficiency, productivity and service, healthcare organizations will soon realize the need to use data mining as an essential tool for meeting a multiplicity of needs.

Needs of the Many Versus the Few

To address the needs of healthcare organizations, effective business intelligence must encompass a broad approach that meets the requirements of multiple types of data users. For example, the chief medical officer responsible for systemwide quality will be interested in enterprise reporting. The infection control nurse responsible for monitoring nosocomial infection incidence requires ad hoc query and analytic capability not only to the departmental level, but also down to the unit and even room level.

The financial analyst requires regular analytics on resource consumption from throughout the organization. The front line physician increasingly requires decision-support based on care-delivery patterns for individual patients. The data warehouse that does not respond to each of these separate

constituencies contemporaneously will not provide sufficient value to the organization. As a result, a comprehensive data model that captures primary data—not secondary, derivative data—will become the norm.

As these systems are adopted, it also will become obvious that an analysis of limited data sets (e.g. hospitalization data), while critical to enhancing the quality of care within the institution, will not address the larger questions in healthcare related to the longitudinal impact of care delivery patterns. For example, the total costs associated with deep vein thrombosis precipitated by treatment for a hip fracture cannot be fully evaluated without a complete understanding of the preoperative treatment, post-operative care, and associated rehabilitation, home care and ambulatory services provided to the individual. Therefore, patient-centric models will clearly evolve in support of these data needs.

Furthermore, longitudinal, patient-centric data that is collected to enable data mining and analytics can easily be de-identified to meet various national standards and ease compliance concerns. By de-identifying the data, both intra- and extra-organizational comparative analysis can then be complete, which will greatly contribute to our understanding of care patterns. Benchmarking, protocols, guidelines and other similar quality initiatives can then be built upon a platform of true primary data analytics rather than through the extrapolation from secondary data.

Data-enabled Decisions

The end result is that healthcare organizations will increasingly seek to collaborate at the community level to capture data from all sources: hospitals, physician offices, outpatient care centers, pharmacies and home care programs. While much dialogue and discussion has occurred in recent months on the potential value of the regional health information organizations, the true value will not be in simply sharing data, but in comparing data across organizations. Through such an approach, communitywide standards of care are likely to evolve, which allows for benchmarking at a national level.

As with many other aspects of healthcare, the industry will quickly recognize that basic data capture is not a differentiator in the market place. The actual use of data in the form of altered care delivery patterns can differentiate. Therefore, collaboration on the former and competition on the latter are likely to become the norm over the coming decade. Regardless, the value of data mining will be in assisting users to make better decisions regarding healthcare practices and choices. Providers, purchasers, payers and policymakers alike will benefit from the movement toward effective data analytics in the form of data mining.

Healthcare must embrace more effective data analysis if we are serious about dramatically enhancing the quality of care and promoting patient safety. We must begin to conduct better analysis of care delivery patterns in real world environments using here-and-now data derived from patients in real settings. While understanding that data derived from the point of care can make a difference in our approach to care, dramatic improvements in healthcare will be extracted by combining care delivery data with multiple other data sets.

Continued reliance upon extrapolated, secondary data sources will no longer suffice. Providers must move toward the real-time analysis of data that spans the continuum of care combined with data derived from other traditional sources to support best practices and drive quality throughout the healthcare delivery system.

Possibilities Abound

When effective data mining is deployed across healthcare delivery organizations, we will discover patterns of care that could truly transform the delivery of care. Imagine the Wal-Mart model applied to healthcare. In the not-too-distant future, healthcare analysts will be able to combine data captured from individual patient information from the physician’s office, the outpatient laboratory and pharmacy, and combine it with national weather service predictions, air quality indicators from the Environmental Protection Agency and other miscellaneous data sets to predict—with a high degree of certainty—the increased demand for care among young asthmatic patients between the ages of 10 and 18.

The combination of personalized data from known asthmatics, with ozone and humidity levels, temperature projections and established response patterns of individual patients to specific pharmacological treatment is immense. Asthma is only one example. The degree of specificity that can be derived from data analytics on granular, primary data will truly alter our approach in care delivery by supporting individualized protocols.

It's true that data mining is not at the top of the latest survey of the issues impacting healthcare leaders. However, if the quality-service-cost triad is to ever improve, providers must now commit to deploying systems that offer real solutions derived from useful, actionable information. The power of data mining for improving the quality of patient care is boundless. With accurate analysis, effective interventions can be deployed, outcomes of care enhanced, and overall costs reduced.

Imagine making a difference in the lives of people now—not tomorrow, not after some longitudinal study, not when we get around to it, but now. It's a compelling vision and a necessary goal, and it's within our reach.

Kevin M. Fickenscher, M.D., is Chief Executive Officer for CREO Strategic Solutions, Washington, D.C. Contact him at drkevin@creostrategicsolutions.com

For more information about CREO Strategic Solutions answers for payers and providers, visit www.creostrategicsolutons.com