THE
NATIONAL LAW REVIEW

Big Data in Health Care

Wednesday, September 17, 2014

The use of big data is now widespread in the U.S. health care system. At major industry conferences and in leading publications, big data has become a watchword for clinical care providers, private insurers, pharmaceutical companies, and numerous other stakeholders.

Appreciating the implications of this new era of big data in health care requires an understanding of at least three distinct characteristics of the data itself: (1) the increased quantity of data; (2) the proliferation of new types of data; and (3) the potential of data sets to be assessed in an overlay of one type of information over another to create insights that were unavailable when each of the component data sources was initially compiled.

Each stakeholder includes numerous subgroups. Payers, for example, includes commercial payers and insurers, CMS, and state Medicaid agencies. Manufacturers includes pharmaceutical industry stakeholders, such as pharmaceutical companies and other intermediaries (e.g., pharmacy benefit managers). Similarly, each of the other stakeholders can be further categorized into relevant subgroups.

Article By
Analysis Group Health Care Consulting Services
Analysis Group, Inc.Analysis Group News

Health Law & Managed Care
Communications, Media & Internet
All Federal

Increased quantity of data

Electronic health data are now being collected, coded, and processed by numerous organizations on a massive scale for disparate purposes. The FDA Mini-Sentinel program, for example, which focuses on medical product safety, has access to the health records of more than 125 million people, and Medicaid MAX administrative claims data, which are approximately six terabytes in size, contain billions of historical records of every payment made on behalf of millions of fee-for-service Medicaid beneficiaries in every U.S. state and the District of Columbia. This immense data set has been analyzed in many contexts, including by the Department of Justice in its investigations of manufacturer conduct regarding promotional practices. Longitudinal, patient-specific data of this type provide insights into the extent of off-label prescribing for particular medical conditions and of comorbid conditions and patient drug use history at the moment when the decision to prescribe a drug off-label was made by the physician. A big data set of this type can often be very helpful in identifying larger trends in off-label prescribing and offers a more robust body of evidence than smaller, cross-sectional physician surveys.

> Big Data generally refers to collections of information that are too large to be processed traditionally. This can include data sets that comprise hundreds of terabytes of information (or, at times, even more than that).
>
> Estimates of the total size of big data in health care vary considerably and range into the hundreds of exabytes. One exabyte is equivalent to 1,000 petabytes; one petabyte is equivalent to 1,000 terabytes; and one terabyte is equivalent to 1,000 gigabytes.

Access to this type of large database can also shed new light on the cost-effectiveness of particular health treatments over time. This can be valuable in studies examining rare diseases, specific treatments, or particular patient subgroups, where access to a large overall patient population is necessary to ensure that the sample sizes are sufficient to draw statistically meaningful inferences.

These vast troves of health care information often require a larger scale of raw computing power, such as parallel processing, and new computing approaches that include machine learning procedures and predictive analytics. Health Care Practice Director Paul Greenberg also notes: “Increases in the quantity of available data can change the nature of the analysis required to draw meaningful conclusions. The scope of the data can cross a threshold at which the data become qualitatively quite different from smaller data sets, increasing the importance of human expertise in generating usable results.”

New types of data

As the quantity of available data has increased, information has also grown more diverse, resulting in the proliferation of new types of data that simply did not exist just a few years ago, such as next-generation smart-device biometrics and real-time patient data. Because these data are so new, appreciating the applicability of these information types requires a high degree of industry experience and creativity.

Although some researchers remain uncertain about social media’s compatibility with research requirements, such platforms unquestionably offer new opportunities for health information dissemination that can be leveraged in new ways. These new data could help to track adverse events both from a pharmacovigilance perspective and for drug safety analyses in the context of product fraud litigation. Although many aspects of social media data collection and analysis remain in early stages of development, the use of responsive, real-time information associated with the risk-benefit profiles of drugs is likely to play a larger role in drug safety monitoring over time. This will become particularly important as social media data analysis transitions from straightforward, volume-based assessment to include more sophisticated methodologies that consider cluster effects, herd effects, and source variations across websites to reduce credibility concerns. In a related article, Managing Principal Mei Sheng Duh addresses the use of social media data with respect to pharmacovigilance.

Combined data overlays

The increasing quantity of available data and the promulgation of new data types have also created opportunities to merge multiple, distinct data sets to produce novel outcomes. This additional type of big data could help to clarify potential drivers of prescribing dynamics in the context of various types of disputes. For example, government investigations of alleged manufacturer kickback payments to physicians can often be illuminated by such multilayered data. The insights obtained by overlaying physician prescribing data on speaker honoraria payment histories and event attendance lists can be invaluable in understanding the scope and impact of such events. Although each of these data sets likely was created in isolation, valuable complementarities can be elicited by studying them in combination.

For many years, de-identifed medical and pharmacy insurance claims databases have been a rich source of information for health economics and outcomes research. These databases contain information about patient diagnoses and medical procedures rendered or drugs prescribed. Today, there are opportunities to overlay these data over provider clinical records – such as electronic health records (EHR) and laboratory test results – patient and provider surveys for “quality of life” measurements and treatment adherence patterns, and data generated from new biometric recording technologies to provide a comprehensive perspective on disease severity, treatment, and effects. Such insights could not have been easily obtained prior to the advent of the big data era and the cheap computing power upon which it is built.

> Predictive Analytics
>
> The combination of clinical and financial records across multiple databases makes it possible to track patterns of disease prevalence in entire populations and pioneer new approaches to predictive modeling. It can also magnify existing data concerns (e.g., privacy, quality of care) and liability issues.
>
> Predictive analytic models, some of which are already in use, have the potential to allow clinicians to make medical decisions based on the outputs of algorithms that process EHR systems in real time across multiple health information exchanges. A patient’s medical information can be entered into a predictive analytic system that is designed to identify patient risks and recommend treatments based on available resources.

When almost everything is statistically significant

Today, more than 80 percent of U.S. hospitals have adopted some form of EHR. Within a single hospital, multiple petabytes of data are often divided between structured data (e.g., medication information) and unstructured data (e.g., clinical notes) organized in mixed systems. There is significant variation from hospital to hospital, however, in the successful use of advanced health IT capabilities to improve the quality and cost of care.

Mr. Greenberg explains: “At the scale of big data, formal statistical signals are often superseded by assessments of clinical or economic importance. This is due to the fact that evaluations based on extremely large numbers of observations will likely be statistically significant in most circumstances.” He adds: “This increases the importance of expert judgment in big data analysis. Because health care data are often highly specific to the field – think of medical terminology, practice variation both by region as well as over time, and non-uniform data recording – the ability to draw meaningful conclusions requires analytical expertise supported by a diversity of health outcomes research experience.”