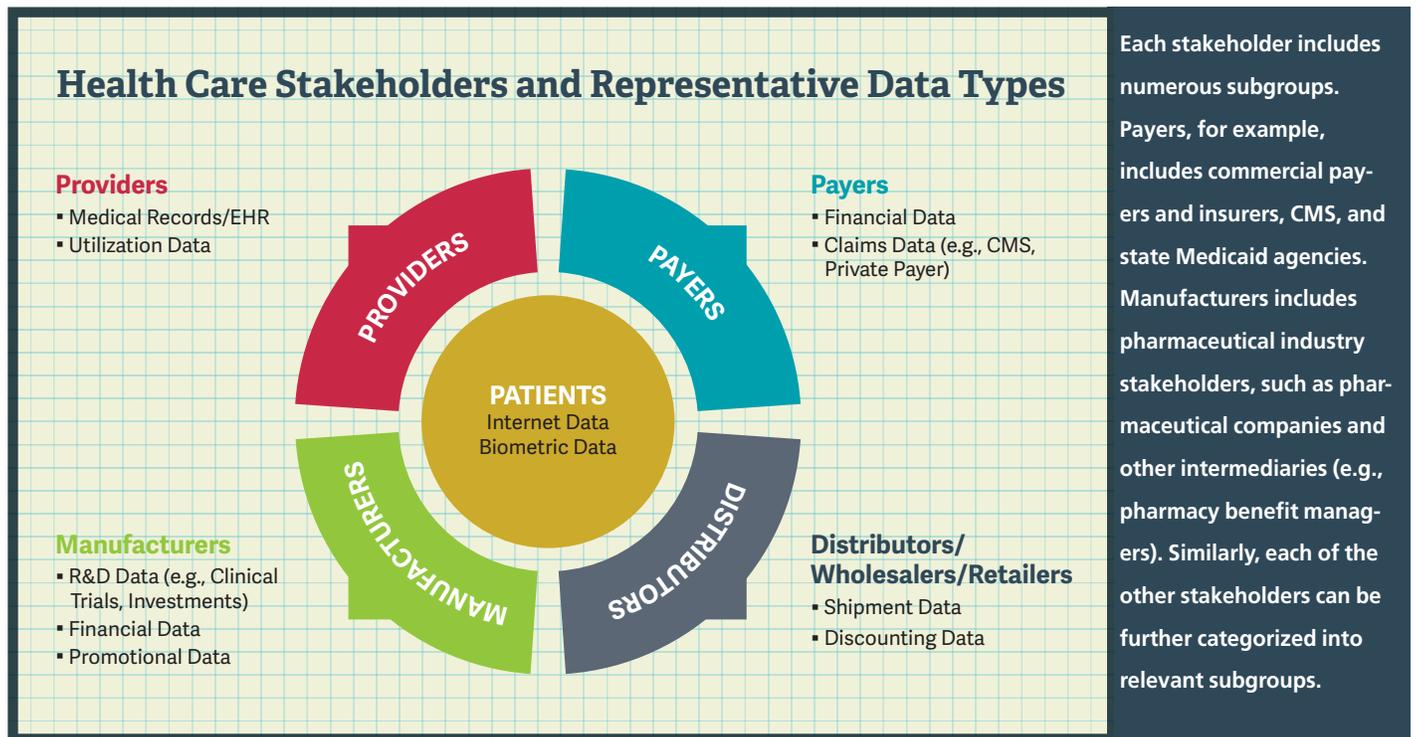


# BIG DATA IN HEALTH CARE

## New Characteristics, Diverse Approaches

FALL 2014 | Analysis Group Health Care Consulting Services Bulletin

**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. In this *Bulletin*, leaders of Analysis Group's Health Care practice discuss aspects of their ongoing work with big data and accompanying analytics, highlighting opportunities to improve decision making and health outcomes research initiatives.



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.

#### 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 (CONTINUED ON PAGE 2)

# Big Data in Health Care (continued from page 1)

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 this issue, Managing Principal **Mei Sheng Duh** addresses the use of social media data with respect to pharmacovigilance (see page 4).

## 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 multi-layered 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.

## Big Data in Health Care (continued from page 2)

For many years, de-identified 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.

### 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, multi-

ple 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.” ■

### 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.

### Web Feature

## ► Diagnostics and Next-Generation Genetic Sequencing

As the costs of next-generation sequencing (NGS) continue to plummet, researchers are able to draw upon increasingly rich sources of data to identify genetic patterns relevant to the risk, cause, and treatment of disease. Next-generation techniques have been introduced in a range of diagnostic applications, including prenatal screening for genetic disease, transplant typing, and risk assessment and treatment selection in oncology. While traditional molecular diagnostic techniques, such as polymerase chain reaction (PCR) and in-situ hybridization (ISH), remain relevant, many emerging applications offer opportunities to draw new associations between the genetic information provided by NGS and disease prognosis or treatment. However, for these advances to generate clinically – and economically – relevant outcomes, researchers face a range of scientific, clinical, regulatory, and reimbursement-related hurdles. In an Analysis Group web feature, Managing Principals **Brian Gorin** and **Edward Tuttle** and Vice President **Matthew Barrett** consider these opportunities and challenges. Read more at [analysisgroup.com](http://analysisgroup.com).

# Monitoring Online Signals: Implications of Internet Data in Pharmacovigilance

The expansion of the Internet has created a variety of new data types related to pharmaceutical drug use and disease areas. Although the implications of this trend for regulators and industry stakeholders are not yet clear, it has become commonplace for Internet users to turn to sites such as Wikipedia in their search for medical and drug information before posting personal information on sites such as Facebook, Twitter, and AskaPatient based on their own experiences.

In recent years, regulatory agencies such as the U.S. Food and Drug Administration (FDA) and the European public-private partnership, Innovative

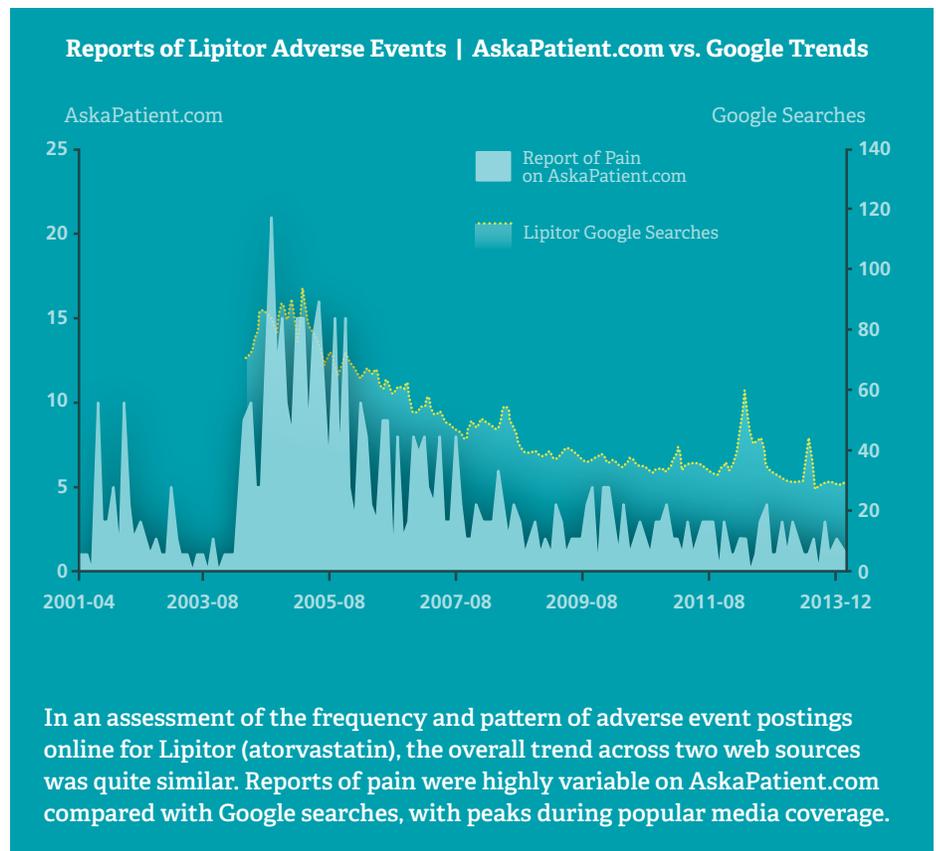
Medicines Initiative, have funded the development of technical frameworks and tools to track and mine social media data for safety signals. While this type of online pharmacovigilance has tended to pinpoint relevant data at the beginning of a drug release and during a major event, it has also triggered concerns from industry stakeholders about potential liability concerns.

Managing Principal **Mei Sheng Duh's** ongoing research in collaboration with several Analysis Group colleagues shows that patients are much more likely to report adverse "quality of life" events online than they are to signal serious clinical events, which are more

frequently reported by physicians. These online signals of non-threatening side effects have the potential to drown out evidence of more serious adverse events, suggesting that self-reported Internet data may more accurately complement, rather than replace, traditional pharmacovigilance data sources. This has significant implications for industry stakeholders. The continued increase in both the quantity and variety of these social media data raises potential issues in health care litigation as well as in health economics and outcomes research (e.g., patient-reported outcomes), accelerating the need for the development of new analytic approaches. ■

As Dr. Duh notes: "Regulators and litigants are increasingly monitoring social media for reports of suspected adverse events, which have the potential to be handled in the same manner as spontaneous reports."

"But big data on social media is often different," she adds. "A spike in adverse signals online could coincide with popular media coverage, for example, rather than with an actual increase in adverse events."



# Addressing Health Care Demand and Big Data in China

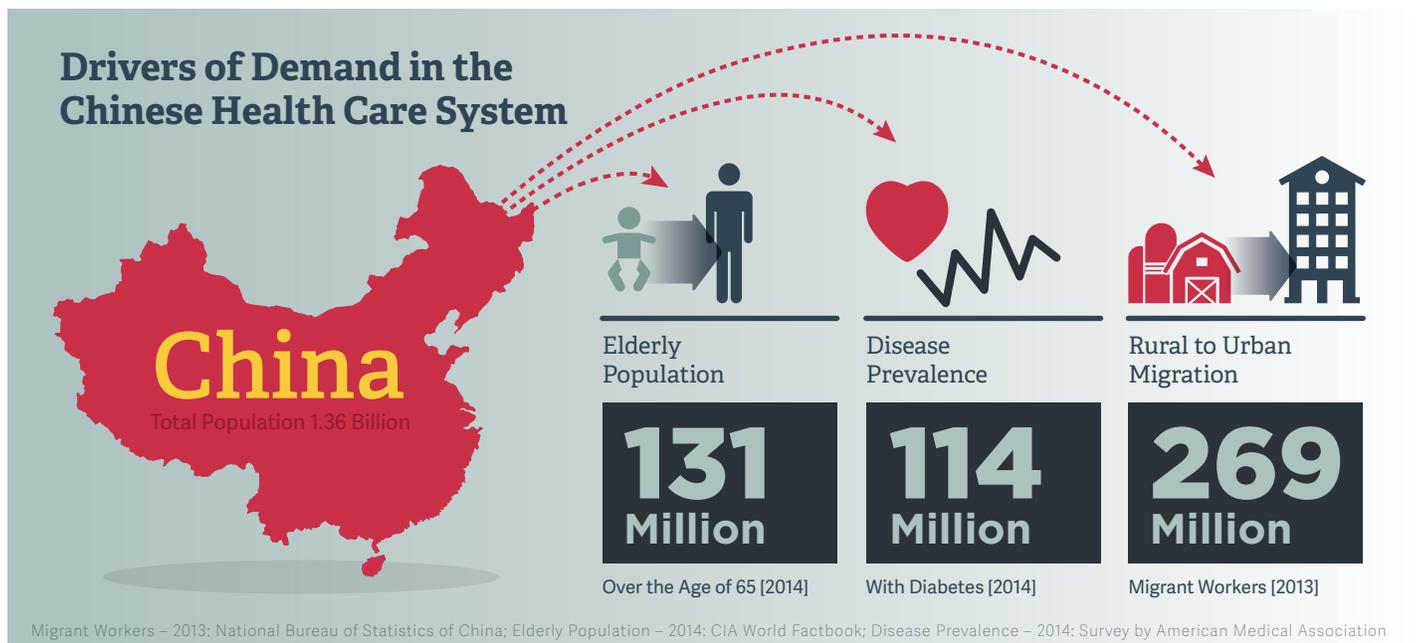
As China's population ages, serious and chronic diseases, such as diabetes, respiratory illness, and cancers, have become increasingly commonplace. China is now home to half of all new patients diagnosed with liver cancer and one-third of all lung cancer and diabetes patients worldwide. This, combined with the continued rural-to-urban migration and rising economic fortunes of the Chinese population, has increased demand for more efficient health care services.

To meet the growing need, health care stakeholders must improve decision making based on data-driven scientific research with respect to the provision of health care resources. For interested researchers, many of the medical institutions now accumulating data across China present a big data conundrum: they have access to detailed data on extremely large numbers of patients, but these data are fragmented and inconsistent as a result of the country's evolving medical care infrastructure. Despite the lack of standardization of data sources, however, the IT infrastructure of the Chinese health care system is reasonably well developed for both providers and payers.

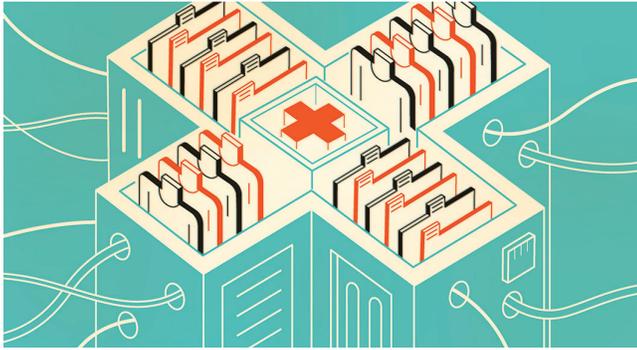
For stakeholders with the ability to navigate China's system, these data represent a significant opportunity. Since 2012, Analysis Group, in coordination with its Chinese subsidiaries and partners, has played an active role in broadening access to Chinese medical data to produce high-quality health care research. Managing Principal **Eric Q. Wu** notes: "Many hospitals in China have already switched from paper records to electronic health records (EHR), and electronically linked databases feature clinical and economic information with large sample sizes, even for rare diseases. Analyzing this data to improve research and health care decision making has the potential to directly impact the lives of millions of people." Dr. Wu adds, "Seizing these opportunities depends on research expertise and experience with the Chinese health care market to manage the consistency, standardization, and reliability of the data." ■

In 2014, Dr. Wu served as guest editor, along with Vice President **Jipan Xie** and Professor Lizheng Shi of Tulane University, of a special issue of *Pharmacoeconomics* focused on China.

In the issue, Dr. Wu, Dr. Xie, and other experts published a study of the treatment pattern and economic outcomes of venous thromboembolism (VTE) among hospitalized patients in China. The study assessed detailed Chinese hospitalization records to better understand patient treatment patterns, including the common usage of conventional anticoagulants and under-monitoring and suboptimal care for patients treated with warfarin, and the high economic burden of care.



# Q&A: A Look Inside Health Care Big Data



Vice President **Mark Gustafson** and **Kevin Brennan**, Manager, Health Care Database Systems, discuss the opportunities and challenges associated with the analysis of big data in health care.

## What makes big data analysis in health care unique?

**Kevin Brennan:** The Affordable Care Act has increased the focus on delivering, measuring, and reporting on health outcomes. This, in turn, has created new requirements for customized information. Due to the falling costs of storage technology, access to patient-specific data that might have once been considered impractical is now a reality. Moreover, other federal regulations also require data collection and access, as in adverse event reporting. The nature of the third-party payer system effectively means that information is managed based on the priorities and interests of the owner, which can make it extremely difficult to develop a unified view of data sources and requirements throughout the system.

## How do you approach such large data sets in your work?

**Mark Gustafson:** The biggest issue is that the data we analyze are a summary of extremely complex interactions between numerous health care stakeholders. In my work with Managing Principals **Bruce Deal** and **Bruce Strombom**, we often encounter data that are not only stored in multiple formats, but are also housed in multiple databases covering different time periods. In a case involving hospital

admissions and transfer patterns, for example, we might analyze electronic health records (EHRs), claim adjudication notes, administrative claims data, and transfer records for hundreds of thousands of individuals. Each type of data set requires different professional expertise (e.g., clinical expertise to assess EHRs, medical and insurance coding expertise to evaluate claims data sets). Many data sets are distinct and were not designed to be analyzed in combination. And don't underestimate the size factor: it's not uncommon for inexperienced analysts to become overwhelmed by the sheer volume of available data, which can lead to poor analysis and unreasonably large time investments. Today, key differentiators have less to do with access to raw computer power than with institutional knowledge about the workings of the industry.

## What capabilities are required for this type of analysis?

**Kevin Brennan:** To integrate heterogeneous data from multiple underlying sources, four interrelated elements are usually required: (1) technical expertise to create appropriate algorithms; (2) technical capacity to analyze and process large amounts of information; (3) human expertise to obtain, understand, and report the appropriate results; and

(4) human experience to appreciate the most efficient and effective approaches, based on similar or related projects. Less technical, but no less important, is the ability to work smoothly with the providers' or payers' technical staff to obtain and understand the data. While this may seem secondary, the client's experience is often shaped by how well consultants interact with their IT departments.

## What opportunities for new health care analyses are enabled by big data?

**Mark Gustafson:** Big data enables the application of new analytical methods. Machine learning procedures can uncover unhypothesized patterns in large data sets and suggest new avenues of research. Predictive analysis techniques enable the development of statistical models that can forecast the likelihood of health outcomes based on treatment patterns. Sophisticated text analytics may be employed to derive analyzable data sets from unstructured data, such as physician clinical notes in an EHR. Careful and thoughtful application of these types of new methods, coupled with appropriate technical and human capabilities, can generate powerful opportunities to address the most challenging health economics questions. ■

## Spotlight

### Analysis Group Experts File Amicus Curiae Brief Regarding the Antitrust Treatment of Pharmaceutical Patent Infringement Settlements

On June 3, 2014, a group of prominent antitrust economists that included Analysis Group Chairman **Bruce Stangle**, Managing Principals **Paul Greenberg** and **Pierre Cremieux**, and Principal **George Kosicki**, and academic affiliates **Henry Grabowski**, **James Hughes**, and **Michael Wohlgenant** filed an amicus curiae – or friend of the court – brief with the U.S. Court of Appeals for the Third Circuit in the Lamictal Direct Purchaser Antitrust Litigation. At issue in this litigation is the proper interpretation of the Supreme Court’s 2013 decision in *FTC v. Actavis*, which established that so-called reverse payment settlements between brand and generic manufacturers should be subject to antitrust scrutiny under the rule of reason, with an important area of inquiry being whether a “large and unjustified” payment flows from the patent holder to the alleged infringer.

The brief addresses the economic principles that lower courts should take into account when addressing the issues raised in the *Actavis* decision. The authors explain that patent settlements may involve consideration flowing from the brand to the generic for reasons other than delay of generic entry and that some settlements may not occur without a payment. The authors also explain that non-cash consideration, such as an agreement by the brand to forgo marketing an authorized generic, should not be viewed with suspicion because it encourages generic entry. Finally, the authors conclude that increased scrutiny of pharmaceutical patent infringement settlements will decrease patent value, discourage innovation, and upset the balance between innovation and generic access that was intended under the Hatch-Waxman Act.

### ISPOR 2014

Members of Analysis Group’s Health Care practice were recognized at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 19th Annual International Meeting in Montreal, Quebec. Managers **Brad Rice** and **Noam Kirson** and Principal **Howard Birnbaum**, as well as Rami Ben-Joseph, Head of Health Outcomes and Pharmacoeconomics at Purdue Pharma, received an award for the best poster at the meeting for their presentation of “The Burden of Opioid Abuse among Commercially-insured Patients.”

In addition, “Cost-effectiveness of Adalimumab for the Treatment of Moderately to Severely Active Ulcerative Colitis in Canada” was selected as a best poster finalist. The poster was presented by a research team that included Manager **Hongbo Yang**. In total, Analysis Group consultants and colleagues delivered one podium presentation and exhibited 27 posters focused on cost-of-illness and comparative effectiveness studies targeting hepatitis C, acute coronary syndrome, multiple sclerosis, sickle cell disease, HIV, hemophilia, schizophrenia, opioid abuse, and lung cancer, as well as health care resource utilization in the United States, Canada, and China.

### On AnalysisGroup.com

▶ **Improper Promotion and Kickback Allegations**

Vice President **Richard Mortimer** considers big data analysis with respect to allegations of off-label promotion or payment of kickback.

▶ **The Depth of Provider Data**

Vice President **Dov Rothman** highlights analytic opportunities created by large quantities of internal hospital personnel data.

## RESEARCH SPOTLIGHT

# Assessing the Economic Burden of Alzheimer's Disease

More than five million Americans are estimated to have Alzheimer's disease (AD), which is the most common form of dementia in the United States. Much of the medical care for patients with AD occurs in primary care settings, but some patients are referred to specialists (e.g., neurologist, psychiatrist, or geriatrician) for further diagnosis and treatment upon the first presentation of symptoms suggesting cognitive decline. While there may be concern that treatment by specialists would result in increased costs of care, new research suggests that consulting with a specialist at the early stages of cognitive decline may result in reduced spending for medical services.

In a retrospective analysis of de-identified administrative claims for more than three million Medicare beneficiaries that included patients who had been diagnosed with AD between 2009–2012, Analysis Group researchers determined that patients seen by a specialist when first diagnosed with cognitive symptoms were more likely to be diagnosed sooner with AD and incurred lower medical costs – particularly in the year following the initial cognitive decline diagnosis.

### Cost Differences for AD Patients Who Consulted a Specialist upon First Presentation of Symptoms



Presenting these findings at the Alzheimer's Association International Conference (AAIC), Manager **Noam Kirson** explained, "These results suggest that seeking timely care from specialists may result in an earlier, more accurate diagnosis and reduce overall medical resource use and costs among patients eventually diagnosed with Alzheimer's disease."

### A LEADING PROVIDER OF ECONOMIC ANALYSIS

Analysis Group provides economic, financial, and business strategy consulting to corporations, government agencies, and law firms. The firm integrates analytical expertise in health care economics with knowledge of industry dynamics, disease states, and business strategy honed from more than 30 years of experience in all aspects of pharmacoeconomics, epidemiology, and health outcomes research. The firm has more than 500 professionals, with U.S. offices in Boston, Chicago, Dallas, Denver, Los Angeles, Menlo Park, New York, San Francisco, and Washington, D.C., and internationally in Montreal, Canada and Beijing, China.

617 425 8000 · [www.analysisgroup.com](http://www.analysisgroup.com)

