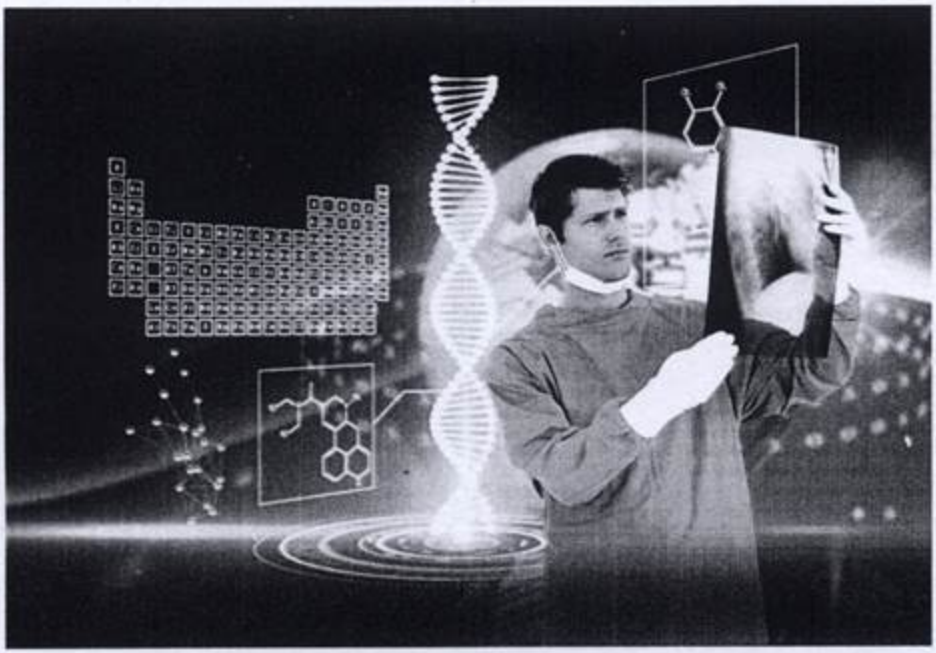
THE DATA DELUGE

There has been a lot of discussion around the potential of big data in pharma and how it could be a game changer in bringing new paradigms to drug discovery, clinical trials and other such areas of the pharmaceutical value chain. By **Shalini Gupta**



Pharmaceutical companies manage vast amounts of information regarding products, customers, prescriber behaviour and clinical trials across hundreds of different types of compounds. However, they are also careful about sharing this information openly, given the risk and uncertainty coupled with getting a blockbuster drug to market. Realising the importance of big data and utilising it sufficiently would be crucial and perhaps the only way forward to getting more value out of it, if experts are to be believed.

Data with benefits

As in sales and marketing, Big Data analytics can be used to manage and analyse data generated with respect to drug formulations, chemical compounds for drugs, details of clinical trials, etc. C Gowri Shankar, Executive Director, TAKE Solutions suggests that rather than creating "value from the volume of data" that is present, it would be better to create insight from it.

Rising R&D costs, lower success rates in the development of new molecules, high expectations from healthcare providers and end customers, shift from single molecule or formulation for the entire population to specific formulations for patients with specific genetic disposition calls for new innovation in drug research. Add to this the rise in the average cost per NME (new molecular entity) or biologic from $2.8 billion (2002 – 06) to $4.8 billion(2007-11).

"A newer approach to drug discovery would require an insight into understanding the efficacy of existing molecules and formulations across different demographic groups and patient types. Big data analysis and insight from hospital patient records (EPRs / EMRs) could go a long way in providing the necessary insights and knowledge," Shankar echoes. This would serve as the starting point for drug development for existing diseases, combination formulations and presently incurable diseases he adds.

This is further aided by the short time cycles for genome sequencing and drastic reduction in the cost of such sequencing. While in 2001 it cost $95 million to read an entire human genome, two leading manufacturers are developing machines that can do so for as little as $1000, in a matter of hours. Big data analysis of all historical discovery initiatives would be a starting point to improve the efficiency of the R&D process. This is likely to be a potential starting point for accelerating development and using old research for newer applications.

Pharma companies are also battling claims as a result of adverse reactions to drugs, post launch. Tracking of adverse events is typically done by a report by the patient to the doctor who then decides after an evaluation if this qualifies as an adverse event to be reported to the company. Multiple formulations that each patient is subjected to, the maturity of the process in each geography, under reporting, lack of patient awareness and ease of access to the doctor further complicate the process. Corrective action is thus liable to be delayed and the scope for minimising the impact of true adverse events on the entire patient population is weak.

However, regulatory changes make it important for companies to take note of adverse events even as social media offers a platform for patients to report such adverse reaction. "Big data analysis for adverse event tracking in the social media could be very

"A newer approach to drug discovery would require an insight into understanding the efficacy of existing molecules and formulations across different demographic groups and patient types. Big data analysis and insight from hospital patient records (EPRs / EMRs) could go a long way in providing the necessary insights and knowledge"

C Gowri Shankar,
Executive Director,
TAKE Solutions

"Starting from patient profiling by identifying the right candidates through analytics of demographic information and historical data, evaluating drug readiness, reviewing previous clinical trial events, intervention through correct drug dosage to remote patient monitoring, big data can help bring innovation and build competitive advantage through the clinical trial process"

Somenath Nag,
Director, ISV & Enterprise Solutions,
ALTEN Calsoft Labs

beneficial to the pharma industry and help provide a heads up before the formal signal detection through the existing systems kick in," opines Shankar.

Somenath Nag, Director, ISV & Enterprise Solutions, ALTEN Calsoft Labs feels that companies can get maximum RoI by using Big data analytics to analyse clinical trial data and genome sequencing data. "Starting from patient profiling by identifying the right candidates through analytics of demographic information and historical data, evaluating drug readiness, reviewing previous clinical trial events, intervention through correct drug dosage to remote patient monitoring, big data can help bring innovation and build competitive advantage through the clinical trial process," he pitches in.

It could also help perform computations of large-scale genomics in a timely and cost-effective manner given that processing of sequence data using traditional BI and data management platforms remains a significant challenge.

Tapping the unknown

However, even as the opportunity remains largely untapped, Shankar puts it best when he orchestrates why companies are gingerly in realising the big data opportunity. "The biggest challenge in implementation of big data projects is that 'we do not know what we do not know'. Companies are wary of investing till they see proof of concepts of possible insights from big data analysis," he says.

The absence of relevant technologies to analyse unstructured data, cost of computing power, lack of viable technology alternatives, and reluctance to reveal confidential data pose other problems. The biggest challenge in big data projects, however, is getting multiple stake holders with tangential objectives to work together especially when they need to commit themselves to resources without an accurate sense of the outcome. These include the sponsor who stands to benefit from the outcome, the owners of the raw and unstructured data, the technology experts, the domain experts and service providers. Innovative models for compensation and monetisation for all stake holders are important in the light of unpredictable outcomes, he adds.

Nag strongly feels that an integrated data-centric approach with a clear owner for each data type across functional silos and through the data life cycle is required for a successful adoption of Big Data initiatives, given that most companies work in silos. A successful Big Data initiative with an increased ability to share data requires rationalising and connecting systems which typically contain heterogeneous and disparate data. Enterprises also have shortage of people capable of developing the technology and analytics platform needed to extract maximum value from the existing data, he adds.

The biggest challenge in big data projects, however, is getting multiple stake holders with tangential objectives to work together especially when they need to commit themselves to resources without an accurate sense of the outcome

GSK has been at the forefront of the big data revolution and along with Bayer, Sanofi, Roche, and Lilly, have collaborated to upload patient-level data collected by clinical trials to a web portal(clinicalstudydatarequest.com), where researchers can submit proposals and request anonymised data from certain studies. Collaborations like those between Optum and the Mayo Clinic where Optum provides analytics tools and data to allow researchers to explore information on 149 million UnitedHealth Group patients, while Mayo contributes clinical data on around five million patients to facilitate research is another model.

Thomson Reuters recently announced the availability of data on more than 180,000 clinical trials through its Cortellis Clinical Trials Intelligence programme, including data on drugs, biologics, biomarkers, diagnostics, and medical devices. The data can be accessed through analytics tools within the Cortellis platform or through an API embedded in other software applications, widening the possibilities for clinical integration and research substantially.

The solution will allow users to discover competitor strategies for specific patient segments, biomarker utilisation, end points and novel insights into disease processes and to uncover vital connections by integrating their internal, proprietary data with the wide variety of reference information available through Cortellis. These are early days for big data in pharma, however, newer models are emerging, and it is likely that as the benefits accumulate and reveal themselves, other stakeholders will follow suit. Ultimately businesses will have to realise the hidden or latent value of all the data lying inert with them.

shalini.g@expressindia.com