REAL WORLD EVIDENCE: A FORM OF BIG DATA, TRANSFORMING HEALTHCARE DATA INTO ACTIONABLE REAL TIME INSIGHTS AND INFORMED BUSINESS DECISIONS

Uttam Kumar Barick1*, Daniel Schwarz2, Behsad Zomorodi3, Arun Gowda1, Rituraj Mohanty1, Martin Komenda2
1 Phamax Analytic Resources Pvt. Ltd, Karnataka, India
2 Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic
3 Phamax AG, Oberwil (BL), Switzerland
* Corresponding author: uttam.barick@fs-researchcenter.com

ABSTRACT — Data has always played an important role in assisting business decisions and overall improvement of a company’s strategies. The introduction of what has come to be named “BIG data” has changed the industry paradigm altogether for a few domains like media, mobility, retail and social. Data from the real world is also considered as BIG data based on its magnitude, sources and the industry’s capacity to handle the same. Although, the healthcare industry has been using real world data for decades, digitization of health records has demonstrated its value to all the stakeholders with a reaffirmation of interest in it. Over time, companies are looking to adopt new technologies in linking these fragmented data for meaningful and actionable insights to demonstrate their value over competition. It has also been noticed that the consequences of not demonstrating the value of data are sometimes leads regulators and payers to be severe. The real challenge though is not in identifying data sets but transforming these data sets into actionable real time insights and business decisions.

Evidence and value development frameworks need to work side by side, harnessing meaningful insights in parallel to product development from early phase to lifecycle management. This should in-turn create evidence and value-based insights for multiple stakeholders across the industry; ultimately supporting the patient as the end user to take informed decisions that impact access to care.

This article attempts to review the current state of affairs in the area of BIG data in pharma OR BIG DIP as it is increasingly being referred to.

BACKGROUND

Data in all its various formats has historically played a significant role in the design of any business model and thus companies are mining data, hoping to increase efficiency and gaining a competitive advantage in the market to outperform their peers [1,2]. This is also true for pharmaceutical firms who for decades have been managing vast amount of data which are critical either in the demonstration of clinical and economic value, in product development or commercialization decisions [3]. However, considering the fact that most of this data involves people or patient level data, incremental costs are being added along with stringent regulations for the generation and handling of such data. In order to minimize the cost of these data while simultaneously maximizing the value, pharma companies are looking to adopt new technologies over time; alongside the struggle to integrate various types and sources of data globally [4]. In order to capitalize the usefulness “It is critical to collaborate with researchers and the technology ecosystem to develop innovative solutions to seemingly intractable problems emerging in healthcare and life sciences today” [5]. Unlike other industries where this might be the norm, the challenge in the pharma industry is the ability to look into the data generated differently and transforming it into scientifically acceptable evidence.

Solution to this may be managing BIG data. The term BIG data is a “blanket term” for any data set which is big in nature and too complex to process using the traditional management tools or data processing applications [6]. The impression of BIG depends on how an organization manages its data. For some companies,
facing hundreds of gigabytes for the first time might imply BIG data while for others it may require 100s of terabytes before considering information as BIG data [7].

BIG data is at the center of many ongoing discussions about data and its next avatar; so much that it attracts a lot of media attention and makes it difficult to differentiate reality from hype. An increasing trend though is the acceptance of BIG data’s importance and integration into growth strategies across all the industries [8], including life sciences and pharmaceuticals.

Concurrently, healthcare stakeholders have started to see the value of evidence based data and its role in optimizing the patient journey or access to care. Over time, payers are becoming skeptic of the incremental benefit demonstrated in controlled settings by new drugs and also the value for money for patients and the economy. Payers are therefore increasingly demanding stronger evidence of value. Stakeholders at a local level have also started investing and developing their own evidence sources to assist decision making. Earlier, data was collected in silos, leading to a fragmented picture of the patient experience, limiting the ability of stakeholders to answer key questions. Now, stakeholders are no longer taking any risk associated with new products/interventions and have started demanding and mining more data to prove value. This marks the beginning of the “Prove It” era and the science of real world evidence has gained importance to bring together disconnected data sources in a meaningful way to answer key questions that various stakeholders pose. Such collaborations are supported by advances in technology; enabling secure and effective ways to link the data sources without compromising the integrity of data. The result – evolution of disconnected data sources into real world evidence leading to real time insights and actions. Real world evidence data sets are often called as BIG data due to their similarities in magnitude and sources [6].

This article attempts to review the current state of affairs in the area of BIG data in pharma or BIG DIP as it is increasingly being referred to.

DISCUSSION

How BIG is BIG DIP (data in pharma)?

The ability to deliver new lifesaving drugs to patients in a timely and cost effective way is dependent on the capacity to manage the huge amount of data generated throughout the phases of a product lifecycle [9]. However, the pharma industry has multiple functional areas and each contributes unique sets of insights to understand the mechanisms of progression of disease. This is making such data management increasingly difficult; pharmaceutical and healthcare firms are managing large data sets from the research records and patient information to utilization details, and supply chain monitoring. Around 70% of a typical pharmaceutical project now involves simply managing data [10] and this involves assimilation, transmission and cleaning before any actual analysis can begin. As per Thomson Reuters’ survey, the biggest opportunity for BIG data in pharma industry is from early-stage discovery (i.e., drug discovery and development) –41.20 percent, followed by understanding the market (i.e., real world) –26.5 percent [11]. Data sources are always associated with the three forms of “V”, i.e., volumes, varieties and velocities of data [12,13]. In the same time veracity and vocabulary are equally important, especially as social media is becoming a vital source for gathering of BIG data, primarily due to self-reported data. With the adaptation of semantic-based methodology [14], RWE studies can facilitate harnessing actionable business insights in real time manner.

What follows are the current analytical and technological trends with examples of transforming RWE to real time insights.

What’s the BIG deal in BIG DIP?

Data mining perspective

The pharmaceutical industry is well known for performing quantitative analyses for clinical and market research. In the marketing departments, data mining (DM) applications are used for sales force planning and direct marketing to doctors and consumers. Data mining techniques are used to a variety of critical business decisions in the pharmaceutical industry [15]. The role of DM methodologies in pharmacological domain is evolving. The field of DM covers varied powerful tools like “association, clustering, segmentation and classification”, which provide better manipulation of the data and help the pharma sector compete on lower costs while improving the quality of drug discovery and delivery methods. “A deep understanding of the knowledge hidden in the pharma data is vital to a competitive position and organizational decision-making” [16]. In general, DM algorithms may significantly assist pharmacological scientists to discover potentially relevant drug-event associations [17]. The obtained results generated by the use of various DM techniques should be viewed as hypothesis proposals and should be evaluated in the context of other relevant data. Alternatively these signals can then be used as a bases for intervention as appropriate [18]. According to Poluzzi et al., the application of data mining techniques in clinical pharmacology can be broadly grouped into two following areas:

- “Identification of new effects of drugs (mostly adverse reactions, but sometimes also new therapeutic effects, and effects in special populations)”;
- “Appropriateness in drug use (e.g., frequency of use in patients with contraindications, concomitant prescriptions of drugs known for the risk of clinically relevant interactions)” [19].
With respect to the increasing trend of volume (amount of data) and variety (range of data types) an application of advanced mining approaches is required. Ranjan [20] presented a set of relevant data sources, which are typical for the pharma industry: rules of general guidelines; clinical data (patient data, pharmaceutical data, medical treatments, length of stay); administrative data (staff skills, overtime, nursing care hours, staff sick leave); financial data (treatment costs, drug costs, staff salaries, accounting, cost-effectiveness studies); and organizational data (room occupation, facilities, equipment). Nevertheless, there is concern about the lack of systematic, objective validation of the methods in this particular area. Unfortunately, a gold standard to validate DM methods does not exist, although various imperfect reference standards may be used to obtain useful insights on the performance of any DM method [16]. The whole process is not only about data analysis; but taking informed decisions from the mined information and fulfillment of the need to communicate analytic outcomes simply and clearly. This is where the practical use of visual analytics approaches, which bring an innovative and effective way to deliver the knowledge from a particular domain to an individual user, is helpful. In general, the power of two robust scientific fields, data mining and visual analytics, can be successfully used to identify and display novel, valid and potentially useful patterns mined from huge databases and warehouses. Moreover data visualization helps in providing high level of understanding and trust to a user not initiated to such techniques [21]. In the most cases it brings transparent and understandable overview on achieved results for different groups of involved stakeholders.

Visual analytics is a new and progressive interdisciplinary field of study that calls for a more structured scientific approach to understanding the effects of interaction with complex graphical displays on human cognitive processes [22]. The higher-level visualization summaries can provide a framework for understanding the immense volumes of data and that reveal unexpected relationships have come to the forefront and it brings the most effective way to present and understand large sets of data [23]. Health care offers many potential applications for visual analytics including pharmaceutical development [24]. A survey presented by G. D. Sun et al. [25] reviews and classifies recent work in this particular domain into a set of application categories in many sectors of human interests and gives evidence about the importance and practical usefulness of visual analytics in real life.

Below the various case studies associated with extraction of unknown and potentially useful information from different areas of pharmaceutical domain are described. Usually DM techniques are used to support the clinicians at the point of care delivery, the controlling of clinical treatment pathways, the administrative and management tasks, and efficient management of organizational and financial data [20]. Parkinson et al. applied advanced informatics and data-mining tools to identify plausible preclinical gastrointestinal (GI) effects that may be associated with nausea and that could be of potential use in its prediction. The objective of this analysis was to assess whether preclinical GI observations and/or measurements can be used to evaluate the risk of drug-induced nausea in man. This work highlighted the usefulness of visual tools for displaying a large amount of data. For example, using a visual network to plot many preclinical effects that are associated with drug-induced nausea in man can help to quickly identify the most relevant findings [26]. Coulter et al. examined the relation between antipsychotic drugs and myocarditis and cardiomyopathy with the use of neural network architecture. They calculated the strength of dependency between a drug and adverse reaction using a logarithmic measure of disproportionality. The final analysis suggested that antipsychotic drugs other than clozapine may be associated with myocarditis and cardiomyopathy [27]. Harpaz et al. showed that a rich and diverse portfolio of data-mining approaches aligned to different strategies and objectives are now available for the analysis and detection of post approval adverse drug events. Published results offer unique prospects that collectively can advance the science of drug safety surveillance [28].

Decision makers in the pharma industry start to recognize the relevance of the definition of drugs and products in relation to management information. In the confusion between costs, care-results and patient satisfaction the right balance is needed and can be found in upcoming effective design and evaluation of graphical information systems that better support cognitive processes in areas as diverse as scientific research and emergency management [16,29].

Answering the basic questions through novel journey of discovery

There is no doubt that BIG data and technology have enabled providing answers to questions that in the past were only answered by primary research. For example the introduction of EMRs (Electronic Medical Records) has helped in the transformation of clinical data into “liquid” usage, and further simplified mining the data for specific queries and needs [30]. This marks the introduction of information technology in the healthcare industry and eventually complements existing technology like business intelligence and data warehousing practices. The advances in the technical capability can help the industry to combine the unstructured clinical and non-clinical (e.g., claims or insurance) data without compromising the patient privacy [30]. EMR’s have changed the way we understand and map the patient journey from diagnosis through adherence of treatment and to the desired or undesired endpoint. This data ranges from baseline assessment at a family physician or GP level, to specialist assessment
Barick UK et al.

at a consultant level, to diagnostic data from the local and/or specialized lab reports, to treatment, and adherence data from claims.

Figure 1 is an attempt to map the patient journey to different sources of real world data that form individual robust databases from various settings of care throughout the progression of disease.

A combination of these datasets can generate evidence based patient journeys by disease, intervention or product that will answers key research questions such as:

- a) Number of patients that have gone through this journey
- b) Number of patients who never get drug treatment
- c) Adherence ratio to the treatment regimen
- d) Patients availing multiple treatments options, and most importantly
- e) Cost burden and its implications on patient behavior [31]

Assessment of a combination of these data sets could potentially outline the cracks in the healthcare delivery system for the providers to fix. The potential of this is immense and the limit seems to be the ability of the human mind to arrive at novel ways of visualizing the hidden information in these datasets.

Transfoming RWE into business insights and actions

Due to rise in the market competition, the pharma industry is under pressure to create and deliver the right treatments at the right time, for the right patients, at the right cost [10]. Therefore, big data in healthcare is venturing beyond the realms of improving profits and reducing overheads and is being used at advanced levels to predict epidemics, cure diseases, improve quality of life and tackle avoidable deaths [32]. Thus pharmaceutical companies need to measure and monitor metrics and trends of all types, faster than ever before. Big data from real world provides the necessary foundation for greater insight in less time, shaping a more effective complementary process of “value” evidence generation both in and outside the clinical drug development process [33]. A well designed study is always based on its foundation of methodology development and setting up of objective segmentation. Implementing a RWE is not much different than implementing any other clinical studies. However; a special attention is required in setting up of the objectives of the study to generate core data, identifying the patient outcomes or endpoints.
Real world evidence: a form of big data, transforming healthcare data into actionable real time insights…

from the define target population. These anonymous data points can be collected through online electronic portal in a real time basis to generate reports which eventually answers the informed business questions.

Figure 2 is the pictorial overview, illustrating-how analytics based decision support can convert real world evidence to real-time insights for informed business decisions in a hospital network setting.

FUTURE PERSPECTIVE

Pharma companies strive to innovate, create and deliver the right treatments at the right time to the right patients at the right cost. RWE is widely foreseen to be one of the instruments towards achieving this. At the same time, RWE could potentially fulfill the demand for evidence from health care decision makers who are continuously seeking value for money along with clinical outcomes involving doctors and their patients. By interlinking healthcare providers, payers, patients and other relevant stakeholders, RWE is likely to become central for establishing effective healthcare frameworks in future.

CONCLUSION

The buzz around BIG data and the hype generated around it is obvious and everyone is interested to acquire this new art of data understanding. However, evidence and value development frameworks need to work side by side, harnessing meaningful insights in parallel to product development from early phase to life-cycle management. This in turn creates value-based insights with robust foundational evidence for informed business decisions, post launch coverage and reimbursement evaluations.

According to Toby Leete and Dr. Ekta Sood, the ultimate utility of RWE is in improving the R&D cycle so that we can get safer drugs into the hands of patients quicker to allow them to achieve better outcomes [34]. Going forward there will be an increasing pressure on pharma companies in terms of reducing the cost of treatment with the demonstration of real value and in this situation “RWE will be vital throughout the life-cycle of every product” [35].

BIG DIP in the form of numerous datasets is omnipresent due to the nature of the healthcare industry in general. The real challenge is not the lack of data, but the awareness of where this data exists and establishing frameworks to mine these. The varied nature and size of these datasets makes it all the more important to be able to convert the data into visually appealing and digestible formats. These should in-turn facilitate the ability of pharma managers, R&D, payers, physicians and ultimately patients themselves to take informed decisions that impact patient’s access to care. Caution however needs to be applied with each step as again the nature of these data mean there is
confidential information involved and the integrity of data has to be maintained. Whether BIG DIP derived from the real world evidence dispersed across the health delivery ecosystem will be able to revolutionize the time to discovery and access to a drug or intervention remains to be seen. But irrespective of how this evolves we are in the eye of the storm when it comes to transformation in the pharma industry from reliance on structured controlled data to opening up to unconventional yet valuable datasets.

EXECUTIVE SUMMARY

BIG data in pharmaceutical industry

The historical importance of data in the design of any business model is immense and thus companies are mining data to increase efficiency and gaining a competitive advantage in the market. Unlike other industries where this might be the norm, the challenge in the pharma industry is the ability to look into the data generated differently and transforming it into scientifically acceptable evidence.

It is believed that most of this data involves people or patient level data that requires special measurement to ensure the privacy of the patient. In addition to this, there are multiple functional areas within the industry that contributes unique sets of insights to understand the mechanisms of progression of disease. This is making data management increasingly difficult. Based on the volumes, varieties and velocities, pharma data sets are termed as BIG DIP.

Dealing BIG DIP in perspective of data mining

The role of data mining methodologies in pharma domain is evolving. The field covers varied powerful tools which provide better manipulation of the data and help the pharma sector compete on lowering the overall costs of the treatment while improving the quality of drug discovery and delivery methods. DM techniques are also used to support the clinicians at the point of care delivery, controlling of clinical treatment pathways and efficient management of organizational and financial data. Visual analytics is a new interdisciplinary field of study which provides a framework for understanding the large volumes of data and it brings the most effective way to present and understand large sets of data.

Answering the basic questions through novel journey of discovery

The introduction of electronic medical records has helped in the transformation of clinical data into “liquid” usage, and further simplified mining the data for specific queries and needs. EMR’s have changed the way we understand and map the patient journey from diagnosis through adherence of treatment and to the desired or undesired endpoint. On integration of these data sets can generate evidence based patient journeys by disease, intervention or product that will answers key research questions.

Transforming RWE into business insights and actions

BIG data from real world provides the necessary foundation for greater insight in less time, shaping a more effective complementary process of “value” evidence generation’ both in and outside the clinical drug development process.

BIG DIP in the form of numerous datasets is omnipresent due to the nature of the healthcare industry in general. The real challenge is not the lack of data, but the awareness of where this data exists and establishing frameworks to mine these and convert the data into visually appealing and digestible formats.

REFERENCES

Real world evidence: a form of big data, transforming healthcare data into actionable real time insights...


