



Clinical Research - A Big Data Science Approach

Dr. Elton Mathias, MD MPH

Director, Clinical Research,
Innvocept Solutions, Mumbai,
Maharashtra - 400607, India

Dr. Roveena Goveas, MD MS

Associate Program Director,
Internal Medicine Residency
Program, Mountainside Medical
Center, Montclair, NJ 07042, USA

Manish Rajak

Head- Clinical Research Services
Innvocept Solutions, Mumbai,
Maharashtra - 400607, India

ABSTRACT

“Big data” is defined as the collection of large and complex datasets available in structured, semi-structured, and unstructured form which are difficult to process using traditional database management tools or data processing applications. Big Data is also defined from 5Vs which refers to Volume, Variety, Velocity, Veracity and Value. Clinical Research is one of the most important as well as promising part of health research. This article provides an overview on some specific aspects of clinical research when adapted to Big Data Science path ways which could be utilized for transforming millions of data points into predictions & simulation to provide cost effective medicines in reduced timelines. This article also highlights the opportunities & challenges that Big Data brings with it.

Keywords: *Big Data, Clinical Research, Data Science, Clinical Trials*

INTRODUCTION:

Big data is defined as a huge and lots of information available either in controlled or uncontrolled environment. However, it is also refers to the collection of large and complex datasets available in structured, semi-structured, and unstructured form which are difficult to process using traditional

database management tools or data processing applications[1]. Big Data science is considered as an emerging field and discipline which could be one of the most valuable assets not only in the life sciences such as medical and healthcare, but also other domains including educational standards, government prospective, social sciences, financial industry and business opportunities[2-12].

Big Data is also defined from the 5Vs – that represents Volume (Quantitative), Variety (Diversified), Velocity (Generated Quickly in Real Time), Veracity (Trustworthy, Validated, and Accurate) and Value (Relevant knowledge content of the data). Currently, the existing approach or methods being used in Clinical research ; especially conducting Clinical Trials across globe is getting a costly affair with long waiting period for a drug to hit the market. This writing features the recent progress and future advances about the Big Data Impact in Clinical trials/research.

The Big Data Science Impact in Clinical Trials:-

A data survey was carried out by SCORR Marketing and *Applied Clinical Trials in October 16* [13] which indicated that Big data is a crucial element in the clinical research enterprise. The results of the survey have been described in (Figure 1).

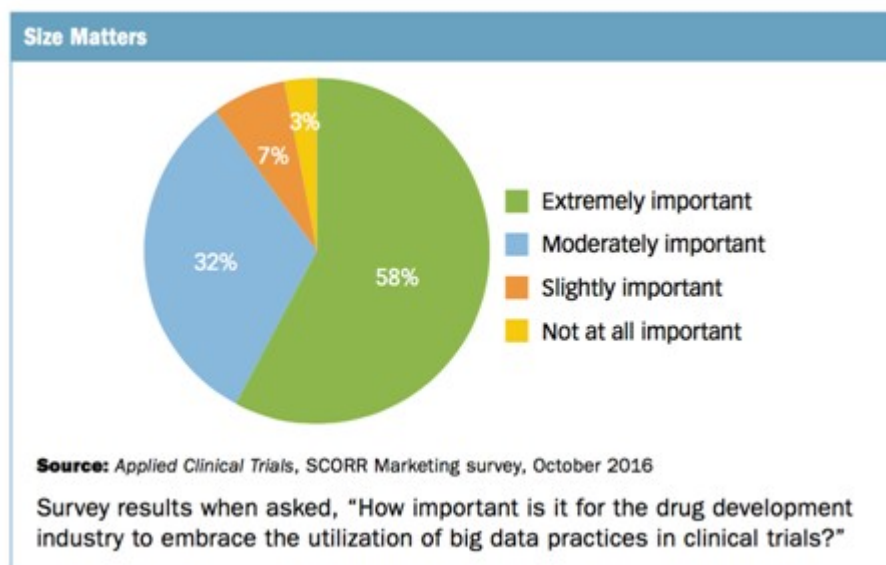


Figure 1:- Survey Results on utilization of Big Data Implementation in Clinical trials

Data is the key element for any research and similarly for Clinical Trials. All new or innovator drugs or medical devices have to undergo efficacy and safety testing in Clinical settings which takes almost 15 years and cost millions to bring it to market. Clinical testing are done at different levels according to different Phases (1/2/3/4) and what finally we get is – “Large Quantum of Data” which can be of three types as below [14]:

- Clinical data, which is collected as part of a clinical trial.
- Operational data, which is used to run the business by tracking individual projects, manage operations, measure quality control, and financial results and so on.
- And data captured from paper documents, which are becoming digitized more and more.

From pre-clinical phase to commercial processing, the promise of connecting & collaborating data is not the only key to optimised & better results having a direct impact on pipelines, but also on cost incurred in the industry as a whole. As the McKinsey & Company predicts, applying Big Data to decision making could work to generate up to \$100 billion in added value on an annual basis [15], which is an incredible amount of savings, especially when it is targeted and applied throughout the supply chain.

Today, all companies are moving towards adopting the Big Data Analytical Theory to minimise the cost and foster clinical trial process which will accelerate the transformation of research into therapy.

The utmost factor is ‘**Recruitment & Retention of Subjects**’ as identification of required subjects is very difficult which delays the recruitment timelines and increases the trial period.

Recruitment & Retention of patients throughout the life of a clinical trial is essential in obtaining the best data sets for analysis and subsequent filings. In order to optimize both recruitment and retention, Big Data approach based on evidence, data and a set of tools are being adopted.

- **Organization of Unstructured Data:** -Currently huge quantum of data in many different forms like Patients records (Paper & Electronic), Patient Lab reports, Imaging Reports, Real time data from Medical & Consumer devices, Health Surveys & Insurance Claims lies in a scattered way. These scattered data can be stored in multiple repositories after appropriate categorisation of data. Each and every repository shall contain a tool-set optimized for the managing data domain and preparing these data for high-speed analysis along with optimized storage of that data, both in a structured and unstructured representation. The above stored data then can be used together to perform high level analysis, patient data visualization and used cases analysis to accelerate clinical research and improve patient outcomes.[16]

- **Data-Driven Patient Selection& Recruitment:** Companies are looking for optimized set of tools that can help identify the right patient pool and

target the right sites and patients, ultimately increasing the chances of a successful and expedited trial completion. Patient Recruitment based on data-driven approaches such as data from electronic health records, patient databases containing anonymized health data, demographic and epidemiological data, historical clinical trial data and secondary data are increasingly being followed. Many companies like Quintiles (now IQVIA), Janssen, Shire Pharmaceuticals, Inventive Health(now Syneos Health) &AdherisHealth have started using big data tools.[17]

- **Real-Time Monitoring:** Many big CROs and Pharma companies have started to monitor real-time data from trials to identify safety or operational risks and try to find solutions in real time. Risk Based Monitoring (RBM) is the current method being used as centralized approach to data collection as well as monitoring, leaving behind the traditional approach of monitoring clinical trials which had frequent physical visits to the site and 100% SDV (Source Data Verification)[18].
- **Drug Safety Management:** Current information technologies in this decade has made easy, the storage of data collected on various adverse drug reactions either in structured or un-structured form. Big Data has become the foundation in Drug Safety management for integrating and analyzing the diversified vast information. Efforts are being taken to analyze the pattern of adverse events, from the data extracted through ClinicalTrials.gov which has been maintained in a database for mining, predicting, and visualizing AEs. Drug-AE relationships were extracted from 8,161 clinical trials, where more than 3 million individuals participated. A total of 1,248 drugs and a total of 31,267 AEs were extracted from these trials. The AEs extracted from these trials span across 26 AE categories.[19]

The Big Data Challenges and Opportunities:-

It has been always observed that arrival of any new technology brings along with it some challenges or limitations. Similarly, Big Data also has challenges and some shortcomings as below listed [20]:-

- Evidence of practical benefits of big data analytics is scarce

- Methodological issues, such as data quality, data inconsistency and instability, limitations of observational studies, validation, analytical issues, and legal issues exist.

A total of 3 searches were performed for publications between January 1, 2010 and January 1, 2016 (PubMed/MEDLINE, CINAHL, and Google Scholar) [21] and an assessment was made on content germane to big data in health care. Top challenges that emerged were issues of data structure, security, data standardization, storage and transfers, and managerial skills such as data governance. However, along with challenges, there were also some opportunity cited such as quality improvement, population management and health, early detection of disease, data quality, structure, and accessibility, patient-centric health care, enhancing personalized medicine, improved decision making, and cost reduction.

Conclusion:

Big Data and its potential in health care sector is a revolution in improving the healthcare research facilities to ensure availability of novel medicines to people in a shorter timelines with reduce cost than the earlier methods and this is the need of the hour. The applications of Big Data analysis in the healthcare industry will be more and more widely used in the future.

Literature review has revealed that though big data has its own benefits & challenges, the benefits overweigh its challenges. Big data analytics need to be integrated into clinical practice to reap the substantial benefits, and clinical integration requires the validation of clinical utility of big data analytics. Today, people need the best quality of life by having the best, quick medicines and health care facilities around them, which Big Data along with the Real World Evidence data can achieve.

References:

1. D B Archaya ,Kausar Ahmed (2016)A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools.IJACSA:Vol. 7, No. 2
2. Murdoch TB1, Detsky AS (2013)The inevitable application of big data to health care. JAMA 309: 1351-1352.
3. Bates DW1, Saria S2, Ohno-Machado L3, Shah A4, Escobar G5 (2014) Big data in health care:

- using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 33: 1123-1131.
4. Gardner E (2013) The HIT approaches to big data. *Health Data Manag* 21: 34, 36, 38 passim.
 5. Jain SH1, Rosenblatt M2, Duke J3 (2014) Is big data the new frontier for academic-industry collaboration? *JAMA* 311: 2171-2172.
 6. Kvochko E (2012) Four Ways to talk About Big Data (Information Communication Technologies for Development Series).
 7. Ohm P (2012) Don't Build a Database of Ruin. *Harvard Business Review*.
 8. Howe D1, Costanzo M, Fey P, Gojobori T, Hannick L, et al. (2008) Big data: The future of biocuration. *Nature* 455: 47-50.
 9. Helles R, Jensen KB (2013) Introduction to the special issue: Making data -Big data and beyond.
 10. Jacobs A (2009) The Pathologies of Big Data. *ACM Queue*.
 11. National Science Foundation (NSF) (2012) NSF Leads Federal Efforts In Big Data.
 12. Wills MJ (2014) Decisions through data: analytics in healthcare. *J Healthc Manag* 59: 254-262.
 13. Lisa Henderson (2016) How Important is Big Data? *Applied Clinical Trials*. Volume 25, Issue 12
 14. Mark Feffer (2016) The Big Data behind Big Drug Development; <https://insights.dice.com/2016/08/15/big-data-behind-big-drug-development/>
 15. Jamie Cattell, Sastry Chilukuri, Michael Levy (October 13) How Big Data Can Revolutionize Pharmaceutical R&D, McKinsey Center for Government
 16. Applied Clinical Trials Editors (2016) The Role of Big Data in Clinical Trials. *Applied Clinical Trials*; Volume 25, Issue 12
 17. Denise Myshko (2014) Data-Driven Patient Recruitment. *PharmaVOICE*
 18. Jaya Kumawat (2017) Risk Based Monitoring: What is it and How Can You Benefit
 19. Callie Federer (2016) Big Data Mining and Adverse Event Pattern Analysis in Clinical Drug Trials; Vol. 14 No. 10
 20. ChoongHo Lee and Hyung-Jin Yoon (2017) Medical big data: promise and challenges; *Kidney Res Clin Pract* 36(1);
 21. Clemens Scott Kruse, Rishi Goswamy, Yesha Raval, Sarah Marawi, (2016) Challenges and Opportunities of Big Data in Health Care: A Systematic Review; *JMIR Med Inform*. 2016 Oct-Dec; 4(4): e38