

A Visualization approach for Clinical Trials with Digital Data

Siddharth Kumar Lokineni, Syneos Health, Cary, USA

Vara Prasad Reddy Sakampally, Syneos Health, Cary, USA

ABSTRACT

With billions of dollars and countless hours of effort spent each year validating new drugs; devices; and medical interventions; it's little surprise that the Pharmaceutical industry is turning to advanced technology to streamline the process. Big Data takes center stage in the current Digital World and is creating opportunities for better healthcare. The amalgamation of consumer data from wearable devices and the digital capabilities of the technology in hand provide numerous opportunities in clinical trial process. The interaction between new sources of digital data and best of old methods in clinical trials will unravel new opportunities to researchers. This paper differentiates traditional trials and modern trials approach with digital devices embedded. Further; this paper explores the strategy to integrate structured and unstructured data and integrate this to Big Data framework. Finally; this paper will show visualizations possible on this integrated data to identify hidden patterns and gain better clinical insights.

INTRODUCTION

The term "Big Data" refers to data that is so large, fast or complex and is difficult or impossible to process using traditional methods. The act of accessing and storing large amounts of information for analytics has been around a long time. But the concept of Big Data gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of Big Data as the three V's:

VOLUME

Organizations collect data from a variety of sources, including business transactions, smart (IoT) devices, industrial equipment, videos, social media and more. In the past, storing it would have been a problem – but cheaper storage on platforms like data lakes and Hadoop have eased the burden

VELOCITY

With the growth in the Internet of Things, data streams in to businesses at an unprecedented speed and must be handled in a timely manner. RFID tags, sensors and smart meters are driving the need to deal with these torrents of data in near-real time.

VARIETY

Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data and financial transactions.

Big Data, irrespective of the industry has been disruptive with the impact. For clinical trials research, Big Data provides the following possibilities – better patient recruitment and engagement to more efficient trials with better quality results, analyzing patient demographics and medical histories to optimizing drug launches, identifying physician behavior and establishing their likelihood to adopt new drugs. Given the many sources from which this data is generated, it's inevitable that trial sponsors and managers will need to partner with third parties to access this vital resource. The pharmaceutical industry is constantly looking to adapt itself to the latest technology to streamline and better the processes involved in the clinical trials.

Most of the pharmaceutical companies and CRO's have adopted technology solutions, including electronic data capture (EDC) solutions and CTMS to make the entire clinical trial process more efficient, financially viable and better the experience of the subjects. In the past era data was mostly collected on the paper, thus making the entire process very complex and prone to human error – this has been taken off with the advent of the electronic data capture and clinical trial management system. Besides this there are a few issues that still remain — particularly, recruitment and retention of patients.

PATIENT FOCUSED MODERN CLINICAL TRIALS

Clinical trial data is undergoing change with the data from health and medical devices. These new data sources are very different from traditional data and are evolving by day. As per the national center for biotechnology information, there are over 318,000 health applications on various platforms and the number of connected wearable devices worldwide is expected to jump from 526 million in 2016 to over 1.1 billion in 2022, while the telecom technology changes from 4G to 5G¹. The value of these Digital Health tools to human health and the healthcare system is still evolving. Over 55% of the most downloaded health apps now use sensor data, Sensor data is the output of a device

that detects and responds to some type of input from the physical environment. Digital sensors linked to health apps are bringing innovation in three key areas: the creation of smart devices, digital diagnostics, and new human-centered clinical trial designs.

The utilization of devices in clinical trials has been standard technique for a long time: syringes, spoons and other simple gadgets have always been there, yet the rate of advancement and innovation has made it hard for the organizations to keep up. It has got to a point where it is practically unthinkable for pharmaceutical organizations to stay aware of the daily advancements. These devices are more advanced than any gadget a pharmaceutical organization would have the option to deliver, as this requires cutting edge innovation skills and aptitude. These gadgets are equipped for checking: pulse, weight, food consumption, sleep patterns, glucose, skin condition, dampness and considerably much more. The following are the challenges today and how the advancement in the technology has improved clinical trial processes -

PATIENT RETENTION

This Pharmaceutical companies have now started asking themselves about the patient dropout rates and what they can do to improve the retention of patients. Integration of latest data collection methodologies will certainly help this cause. Technology will capture the patient data instead of patient providing the data by sticking to a schedule. The advent of electronic clinical outcomes assessment and electronic patient reported outcome solutions for data collection have helped to improve the patient retention. This is because the electronic devices can be used to make the patients work effortless and accurate collection of data and questionnaire process is made easy with electronic patient reported outcomes.

The use of wearables also allows to collect data without requiring patients to visit. Smart watch serves this purpose very well, it has received FDA approval in 2018 for two applications — an EKG and a pulse monitor. Remote-monitoring medical devices can also help patients stick to their prescription schedule conveniently. Examples of such devices that are on the verge of getting in to market include smart pills that can release drugs and transmit data. Another such device is as smart bottle that helps in tracking usage. Technology and mobile providers are offering services such as data feeds, tracking, and analysis to complement the benefits of medical devices.

DRUG CANDIDATE SELECTION

Pharmaceutical companies aim to invest in compounds that are likely to succeed in clinical trials and the market so they can recover initial R&D investments. One of the crucial steps in drug development process is the determination of the most ideal clinical subject who can provide the best probability in regards to success of clinical trials. For this purpose studies are conducted in controlled environment to allow further narrow down the selection of the candidates with desirable absorption, distribution, metabolism, excretion, toxicology properties as well as pharmacology and efficacy to assist in the prediction of clinical 'drug-like' properties.

Here's where electronic patient hospital data and Big Data can help. With analytics and data scientists, patients can be enrolled based on sources other than doctors' visitors, like social media. The criteria for patient selection can now include factors like genetic information, disease status, and individual characteristics, enabling trials to be smaller, shorter, and less expensive.

From the patients perspective there are several different applications that are created that search through a database of various clinical trials and allows patients to find a clinical trial study that is looking for subjects in their demographic. This gives researchers a lot of time to contact the potential subject and explore chances of enrollment. A smartphone application is created by University at Buffalo (Jacobs School of Medicine and Biomedical Sciences at UB) that would allow patients to explore various clinical trials to see if they feel like they would be a good match for the trial. Once the subjects have found the suitable study, they are then able to enroll in it and be a part in the research just by pressing the button as shown below.

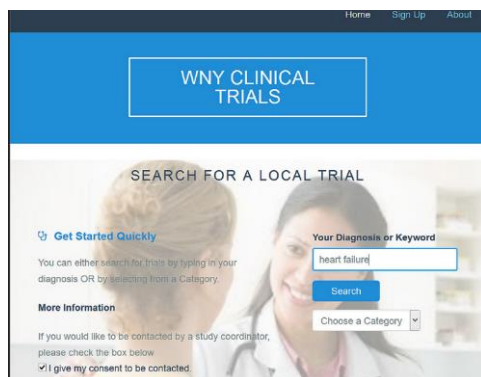


Figure 1. Clinical trial patient enrollment application².

STRUCTURED VS. UNSTRUCTURED DATA AND INTEGRATION INTO BIG DATA FRAMEWORK

Data is heterogeneous, where data is collected in structured, semi-structured and unstructured databases. For the analysis of data, it is necessary to understand these three types of data³:

STRUCTURED DATA

This is considered the most 'traditional' form of data storage. Any data that follows a predefined standard and is ready to be analysed is called structured data. It sticks to a tabular format with identifiable relationships between the rows and columns. Common examples of structured data are Excel files or Datasets. Each of these has structured rows and columns that can be sorted and further analyzed to needs. This makes structured data extremely powerful: it is possible to quickly aggregate data from various locations in the database. In the clinical trials - quantifiable, measurable data such as lab results, blood sugar levels, and cholesterol – can be put to use by clinicians at the point-of-care to aid their decision-making.

UNSTRUCTURED DATA

Unstructured data is information that either does not have a predefined data model or is not organized in a predefined manner. Unstructured information is typically text-heavy and does not follow an easily identifiable pattern. The big problem of pharmaceutical companies is that about 80% of medical data remains unstructured and untapped after it is created. There are two main problems in making this type of data readable.

- The first problem is process-related. Unstructured data is not gathered systematically. An investigator only knows about drug side effects if he or she asks the patient, or if the patient discloses the information. When symptoms are overlooked or patients withhold information, pharmaceutical companies are hindered in their ability to assess the impact of the drug being researched.
- The second problem is technical. Unstructured data is most often recorded on paper in free text or check-the-box fields. Data stored this way makes it very difficult for software systems to interpret, understand, and analyze.

This is where Big Data comes in handy. The ability to analyze unstructured data is particularly relevant when we talk about Big Data, since a large part of data in organizations is unstructured. The ability to extract value from unstructured data is one of the main drivers behind the quick growth of Big Data.

SEMI STRUCTURED DATA

Semi-structured data is a form of structured data that does not conform with the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure.

The reason that this third category exists (between structured and unstructured data) is because semi-structured data is considerably easier to analyze than unstructured data. This reduces the complexity to analyze structured data, compared to complete unstructured data.

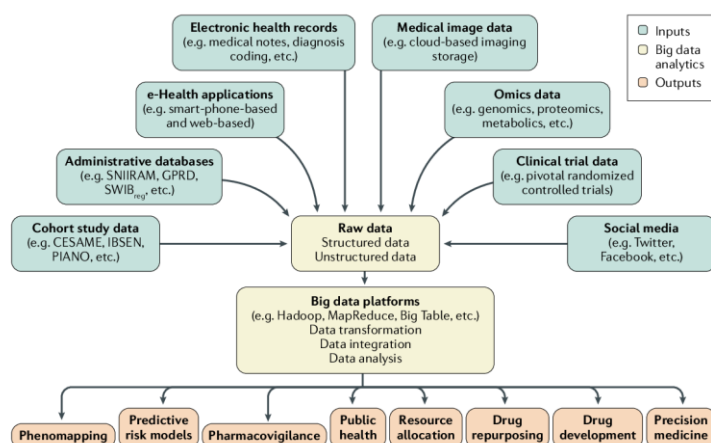


Figure 2 - Framework for data input, processing & visualization in Clinical trials⁴.

The framework shown in the figure 2 – clearly explained us the process flow of the data from various sources in to Big data platforms like Hadoop, MapReduce, Big Table etc. are that provide easy data access of all enterprise data in

a single repository. These data repositories provide fast, responsive access to data to enable real-time operations on various platforms like SAS, Talend, R etc⁵. for data analysis, data integration and transformation. We can make use of these ETL tools for generating reports on our needs in the clinical trials as needed.

BIG DATA AND DATA VISUALIZATIONS

There is so much patient information collected now, and it is very possible to misinterpret it. In order to make accurate decision, it has become very important for the pharmaceutical companies to understand these vast data repositories. Data visualization helps a lot in making allowing the studies to progress in correct directions.

Anscombe's Quartet is a right example of why data visualization is important for better understanding of data. It consists of four datasets presented in plots in Figure 2 below. Simulation data with the same correlation of 0.7 for each case with different sample sizes (200, 300, 400 and 500) and different underlying relationships is used. Dataset 1 (top left) suggests a linear relation and Dataset 2 (top right) a smooth curved relation. In Dataset 3 (bottom left) all but one point lie on an almost straight line and in Dataset 4 (bottom right) all but one point have the same x-value. However, these datasets all share the same mean and variance for x and y, the same correlation between x and y and the same linear regression line. Thus, if we only looked at a table of summary statistics for the below 4 datasets without any graph visualization, the essential differences between them would be impossible to imagine.

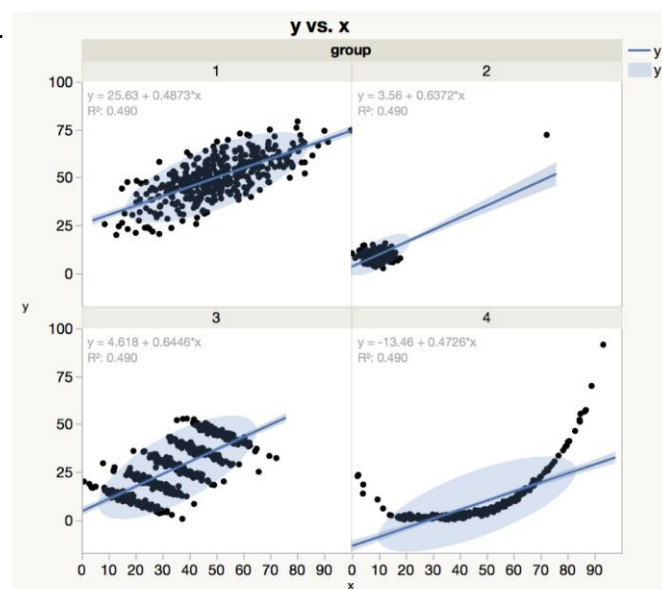


Figure 3: Anscombe's Quartet⁶

Figures are a vital part of any submission to FDA in clinical trials, unfortunately these submissions do not include interactive visualizations. In this blog post we want to present two ideas. Interactive clinical data visualizations can add tremendous value during the conduct of a clinical trial and certain clinical trial data when presented interactively will help in better presentation and provide accurate understanding of the clinical trial data. When utilized data visualization is a powerful tool to investigate data and share detailed insights. It offers a level of detail that normal datasets cannot offer.

The following are the advantages of data visualizations:

- Better monitoring and investigating data,
- Identifying data issues, areas that require special attention, outliers,
- Increasing quality of clinical trial reporting,
- Investigating potential covariates and performing ad-hoc analyses, and
- Sharing and presenting results in greater detail.

A real life application of Big Data visualization is happening today at the Augusta Health, a community hospital located in Virginia, a combination of microbiology data and geospatial analysis is helping to identify patterns in multi-drug resistant organism infections among patients. In this case a cluster of *C. difficile* infections were attributed to a single patient who moved around the hospital as a part of the treatment. The hospital used the hospital floor plans and the time of stay to accurately time and identify the spread of the infections. Using the below visual analytics the hospital could identify hidden patterns in infection rates that would help in altering the way the organization approaches some of its patient safety strategies.

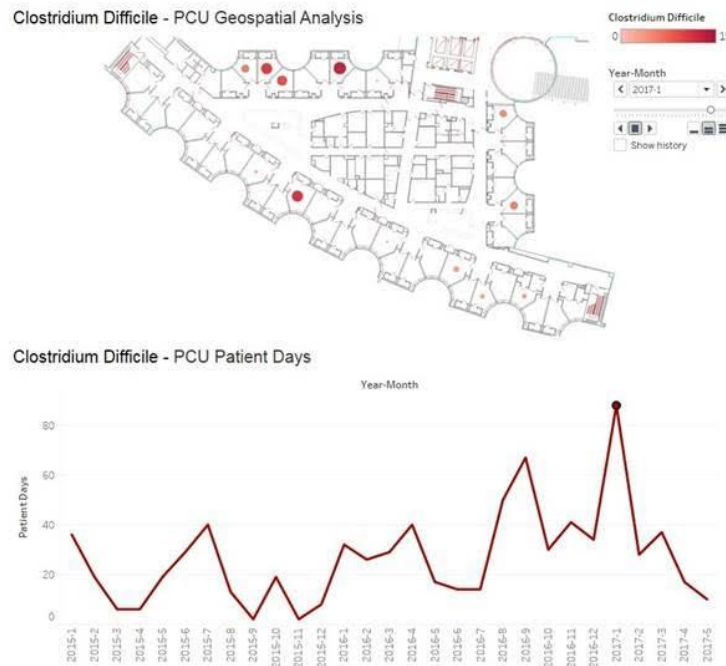


Figure 4: C. Difficile geospatial analysis⁷

This visual analysis of the patient movement gave a better understanding and help to plan a better strategy in the patient safety. Incorporating Big Data and next-generation analytics into clinical and population health research and practice will require thinking, training, and tools. Adequately used, these reservoirs of data can be a practically inexhaustible source of knowledge to fuel a learning health care system. The pharma giants should learn from technology companies. The tech companies are using large, complex data to learn and make better decisions in commercial enterprises. Many companies use immense quantities of information available to them from purchases, searches, and social media to model consumer behavior. Advanced data management and analysis is exemplified by leading information companies, including Google, Facebook, and Amazon⁸.

CONCLUSION

The exploration of big data is a journey of trial and error. There is a strong belief that recognizing the correct pattern is the pathway to the new knowledge bank. The secret of better control over disease prevention and new therapies may be right under the vast ocean of big data, just waiting to be tapped and harvested. The article is not intended to serve as a step-by-step guidance on using big data but to inspire people who are interested in doing exploration on it.

REFERENCES

- ¹<https://www.ncbi.nlm.nih.gov/pmc/articles/pmc5820084/>
- ² <http://www.buffalo.edu/news/releases/2016/06/004.html>
- ³Life in the Digital Vortex, The State of Digital Disruption: 2017. Global Center of Digital Transformation. <https://www.imd.org/globalassets/dbt/docs/digitalvortex2017>
- ⁴ Big data in IBD: a look into the future. Nat Rev Gastroenterol Hepatol 16, 312–321 (2019). Olivera, P., Danese, S., Jay, N. et al. <https://doi.org/10.1038/s41575-019-0102-5>
- ⁵ Augmented Real-time Clinical DataMart http://www.phusewiki.org/wiki/images/9/9d/Augmented_Realtime_Clinical_DataMart.pdf
- ⁶ <https://community.jmp.com/t5/JMP-Academic-Knowledge-Base/A-Bigger-Version-of-the-Anscombe-Quartet/ta-p/22726>
- ⁷Using Visual Analytics, Big Data Dashboards for Healthcare Insights <https://healthitanalytics.com/features/using-visual-analytics-big-data-dashboards-for-healthcare-insights>
- ⁸<https://www.analyticsvidhya.com/blog/2013/07/big-data/>

CONTACT INFORMATION

Author Name: Siddharth Kumar Lokineni
Company: Syneos Health
Address: 3201 Beechleaf Ct
City / Postcode: Raleigh, NC 27604
Work Phone: 210 601 6752
Email: Siddharth.Lokineni@syneoshealth.com

Author Name: Vara Prasad Reddy Sakampally
Company: Syneos Health
Address: 3201 Beechleaf Ct
City / Postcode: Raleigh, NC 27604
Work Phone: 270 282 5166
Email: varaprasadreddy.sakampally@syneoshealth.com

Brand and product names are trademarks of their respective companies.