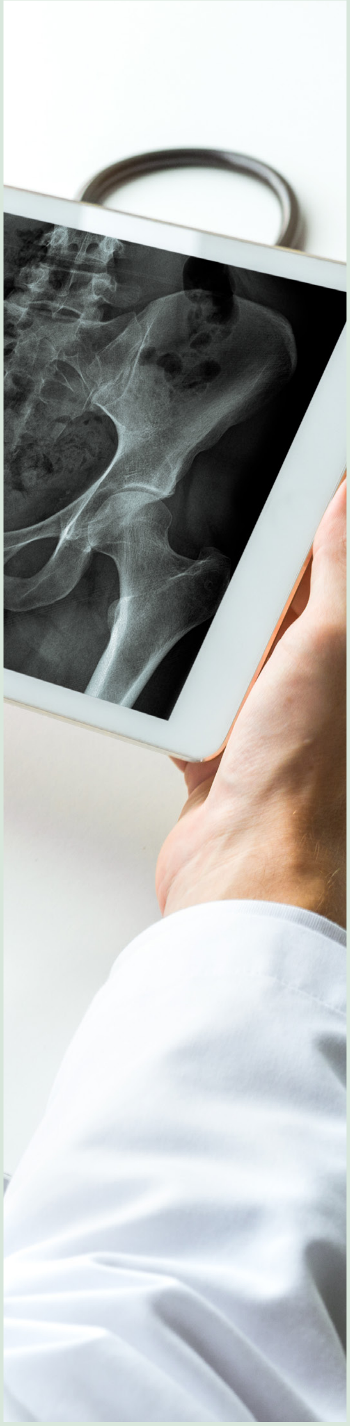
WHITE PAPER



CONNECTING CLINICAL DATA FOR COMPREHENSIVE & INTELLIGENT USE

Infosys®
Navigate your next

Executive Summary

Clinical trials produce a wide variety of data which are stored in secured and controlled systems. Some of them are even forgotten after the trial is over. However, these data hold valuable information that could be reused and can act as a reference if linked accurately. Moreover, it can guide researchers while performing new clinical trials in specific stages, such as – during drug designing, searching for eligibility criteria, assessment schedules, etc. With changing regulatory requirements, and challenging FDA inquiries, having real-world evidence (RWE) which is easily accessible can help pharmaceuticals get their drug approvals faster. In addition to that, RWE can be used in preclinical study design, post-marketing surveillance, etc. Once the data is made available in one place, it will become easier for researchers and clinical staff to access it, leading to faster, cost-efficient, and safer drug discovery which is the need of the hour.

Foreword

"The goal is to turn data into information and information into insights" quotes, Carly Fiorina, former CEO of Hewlett-Packard2. Data is everywhere, but it is of less use unless it is integrated with related entities to derive valuable insights that can be used to make it better, focused, and enable speedy informed decision-making. In healthcare and life sciences, enormous data is generated and collected. Are all these data being used to their full extent?

At nearly every stage of a clinical trial life cycle, from initiation of protocol, statistical analysis plan (SAP), collection of baseline data at participant enrollment, to the analysis of study data generated, documents are available in different formats and are stored in various locations. These data are mostly unstructured, scattered, and hardly shared across industries. These unstructured texts hold an enormous amount of information, that if converted to knowledge, can then be brought to a single platform to get valuable insights.

As of 4 January 2021, there are approximately 362,822 studies that have been carried out across the world in hospitals, universities, doctor's offices, and community clinics which are registered within the https://clinicaltrials.gov/ (registry of clinical trials run by the United States National Library of Medicine at the US National Institutes of Health). Results are posted for about 46,767 studies, out of which 94% are interventional and 6% are non-interventional1. For these studies, clinical data can be available in different forms and formats, such as - documents, databases, images, etc. Enormous diagnostic and lab test data are generated from clinical trials which are currently not being used to the fullest. The question really is - how can we leverage this data to reimagine medicine and allow insights into patients' health? Apart from other information that clinical trial documents hold, the inclusion and exclusion criteria, baseline and study arm, primary outcome, a secondary outcome, and endpoints information provide insights into the reasons behind the execution of the clinical trial, and the success and failure of a study.

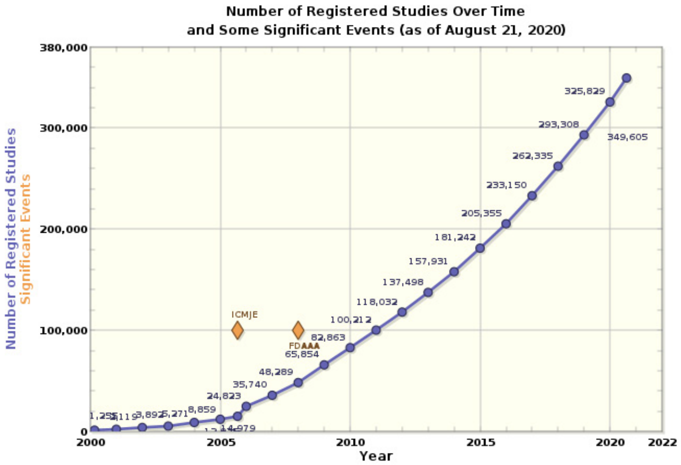


Fig. 1 Number of registered studies by year. Source: https://clinicaltrials.gov/ct2/resources/trends

ICMJE: Indicates when the International Committee of Medical Journal Editors (ICMJE) began requiring trial registration as a condition of publication (September 2005)1

FDAAA: Indicates when the registration requirements of FDAAA began and were implemented on https://clinicaltrials.gov/ (December 2007)1

Clinical trials and patient diagnostic test reports produce huge amounts of beneficial data. An MIT study shows that approval rates ranged from a high of 33.4 percent in vaccines for infectious diseases to 3.4 percent for investigational cancer treatments by inference drawn from diagnostic data4. Additionally, the use of biomarkers to stratify patients tended to be more successful than those that did not, pointing to the growing role of companion diagnostics data in improving clinical trial success rates and future designs4. Linking diagnostic biomarkers data and making them readily available can provide a hassle-free approach for researchers.

Problem statement

Clinical trials produce a large amount of information that is stored in controlled and secured systems. Collecting and storing the data is expensive. Most of the data accumulated are used for single submission and are then forgotten. Researchers and other clinical staff look for this information while setting up a new study or while performing analysis to compare previous endpoints and outcomes. Researchers would also search for a similar set of patient populations to build inclusion-exclusion criteria and study designs while building newer studies. The assessment schedule is another important aspect that links to multitudes of other clinical trial data available in case report forms, summary sheets, etc. of a clinical trial. It is used to understand the assessments performed during a patient visit. Often users such as clinical data scientists and researchers find it difficult to fetch this information because the data is tightly controlled and spread across different systems which means that users must manually trawl through them. Moreover, there is so much variability in data, for example - some are images, pdfs, texts, etc. The need for access to integrated data to search treatments, drug actions, and adverse effects has been particularly felt in the global efforts to address the COVID-19 pandemic. In this case, patient information and other healthcare data such as demographics of affected patients, their treatment, vaccinations, etc. are spread across continents. Integrated data and results of previous studies can help researchers identify possible treatments and adverse effects of drugs quickly. Besides clinical trials, hospital and office visits generate and collect a lot of vital data which is hardly repurposed. Having this data in a format that is dependable and well-linked is one of the challenges that pharmaceutical companies are facing.

Contrastingly, the use of data in commercial industries is more progressive. Amazon has become our everyday example of how data can be collected as users shop; to make everyone's shopping experience easier and more convenient. Once you enter the Amazon site, it can predict and suggest to you things that you wish to buy from simple keyword searches, as well as by showing you similar items that others might have purchased with similar searches or selections. How does Amazon do that? By analyzing every click, selection, search, etc. It has sophisticated tools that keep track of items that you and everyone else have shopped for before. It can then predict things using this data that might appeal to that specific user e.g., offers or discounts on specific products that you have been searching to encourage customers. It collects data holistically from all users, and then analyzes it and links it across similar selected entities. Upon specific user searches, geographies, and known user information, it gives the best/closest match.

While big data and analytics are widely used in commercial and marketing fields, their use in healthcare is limited to a certain extent, mainly because most clinical data are held in organizational, systematic, or regulatory silos. Additional constraints to data access and results of pharmaceutical research are typically not available to wider audiences (internal or external to that organization) until all the necessary steps and approvals are in place.

Content

Clinical data is either collected during ongoing patient care in hospitals or clinics or as part of formal clinical trial programs. They can be broadly classified into the following areas based on where they are collected from: Electronic Health Records (EHR), administrative data, claims data, patient/disease registries, health surveys, and clinical trial data. EHR consists of patient diagnosis information, prescription drugs, laboratory tests, physical evaluation data, hospitalization information, etc., whereas claims information holds billable interactions between insured patients and the healthcare delivery system. This data gives us information on the duration of patient stay in a hospital and costs and can be useful from a healthcare expenditure and economic standpoint. Patient disease registries are specific to certain diseases which require frequent monitoring and care, such as - heart disease, diabetes, and cancer. Clinical information may also be gathered through national health surveys which are more easily accessible since they are collected for research purposes. This information is currently available but in different datasets, and different locations and is not linked. Some of them are also restricted through individual agreements in private companies.

Data from clinical trials, hospitals, clinics, and measurements from sensor devices all contribute to real-world evidence (RWE) data. Through technological advancements, devices can now monitor patient health information such as blood glucose level and blood pressure anywhere, even when they are participating in clinical trials. Data generated from this can be used for multiple aspects, for instance comparing data from previous clinical trials to newly collected patient information to study the efficacy and safety of a drug. Primarily the data generated from clinical trials for a study is used for getting drug approvals, understanding adverse drug reactions, patient safety, etc. However, in today's world, FDA enquires are not just limited to that, they seek real-world evidence. This is where secondary use of clinical trial data will continue to be useful other than the drug development itself. Diagnostic and laboratory tests such as blood pressure, urinalysis, and blood chemistry that are performed at different intervals of clinical trials produce useful data. If we had access to this data, and they were normalized for linking together, then it could be used in enhancing knowledge to assist in disease management and improving patient outcomes. An example of such use is the successful implementation of a nationwide database for enabling real-time surveillance of diseases in Denmark3. Existing data from clinical trials can provide pharmaceuticals with convincing and incremental results. Moreover, with the advent of maturing and evolving technologies, such as Natural Language Processing (NLP) and semantic search, one can find the information in a few clicks.

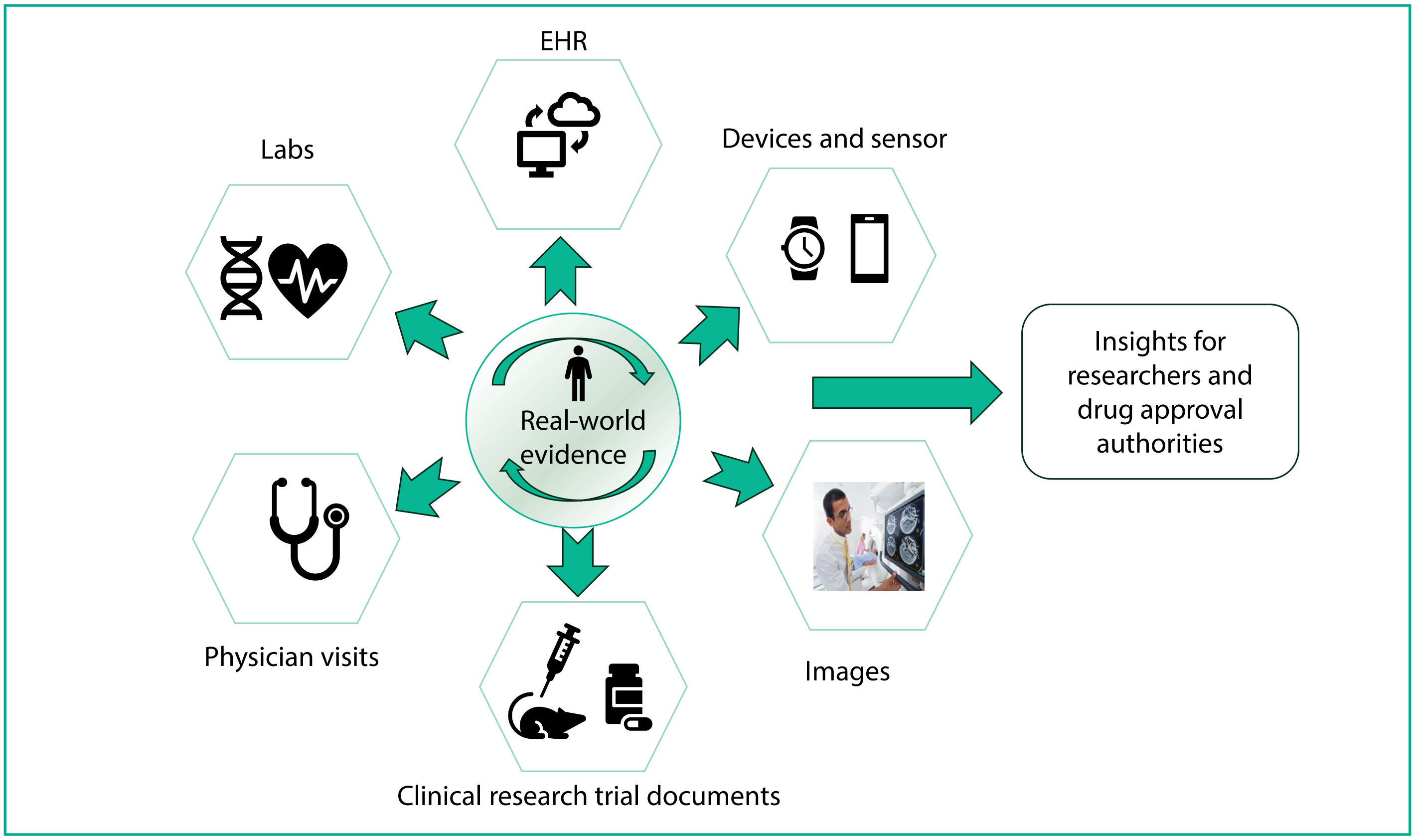

EHR
Labs
Devices and sensor
Real-world evidence
Insights for researchers and drug approval authorities
Physician visits
Images
Clinical research trial documents

Fig 2: Use of real-world evidence information

Here we are talking about data that could be made easily accessible, predictable, and could flow from drug discovery, development, medical affairs, and real-world data sources and used after approvals and even between external partners such as Contract Research Organizations (CROs). With this approach of linking data together for faster searchability as in the case of the Amazon example, researchers, data scientists, and people involved in the clinical trial area would be able to look for information of their interest, categorize it and save it for future reference. For example, consider a researcher who wants to look for all previous studies conducted on diabetic patients on the compound 'Atorvastatin'. They would not only be able to narrow down their search using the keyword 'Atorvastatin' and 'diabetic', but also be able to look for suggestions and insights related to diabetics. This would direct them to all previous studies having Atorvastatin and diabetic mentioned in the protocols. Not only that, wouldn't it be interesting for users to get suggestions on frequent searches? E.g., people who looked for "Diabetic", also found "Cholesterol" search useful. Connecting data not only aims at improving the linking of data elements, but also enhances linkage among all stakeholders in drug research, development, commercialization, and medical affairs to extend their data networks.

Below mentioned are some of the possible challenges that might be encountered while trying to provide a better platform or while attempting to integrate healthcare and clinical information.

1. Initial efforts to find finalized documents from legacy systems.
2. Differences in formats and structure related to images and text symbols expose one to errors within these documents. This proves that every writer has a different approach and without simplification of this data, it cannot be used collaboratively. Nevertheless, this shows that there is an opportunity for improvement and standardization of clinical documents.
3. Difficulty in mapping the data since different authors might have published them. These are required to be standardized or categorized without changing the actual meaning.
4. Language barrier - Health information for certain data such as national surveys may be available in different languages.
5. Heavy reliability on the automation of certain tasks becomes essential while working with huge files.

   Some other common challenges that could be seen are – violations of HIPAA laws since clinical data is involved. This can be resolved by de-identifying and providing role-based access to ensure that specific data elements are visible only to those who are authorized to see them.

Applications

This integrated information has boundless application within the organization as well as for collaboration with external partners, some of which are mentioned below:

- Integrated health care information and having real-world evidence could aid in providing holistic care to patients.
- With an increase in the number of technological applications within pharmaceutical industries, clinical trial data can help data scientists to use stored and real-world data to run analyses and compare results.
- A part of the document could be recreated by incorporating machine learning on top of this extracted data. To take this to another level, this data could be used to compare similar studies and automate the building of clinical trial forms and documents.
- Hints such as participation of patients in previous studies provide insights into where participants might have participated in a successful study, increasing opportunities for participation by the same or similar subjects in upcoming studies. 'Genetic information' is one such criterion that is being sought very frequently when selecting subjects. Proper structuring and availability of information enable trials to be smaller, shorter, less expensive, and more powerful.
- Having essential information in these documents is also useful to answer queries by health authorities in a consistent manner.
- More understanding of adverse events could be gained from adverse events forms or case report forms, which might lead to improved drug delivery and better outcomes for patients.

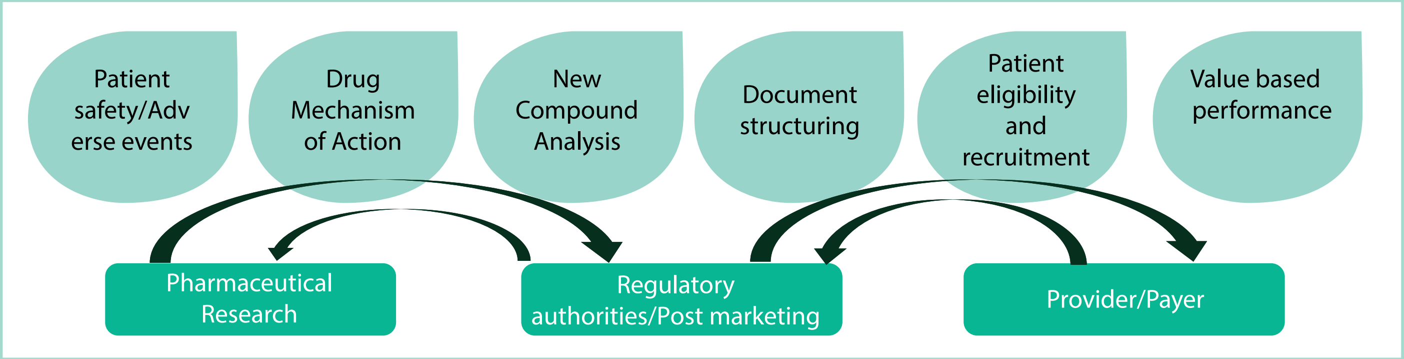


Fig 3: Expansion of clinical data network to other healthcare sectors

Making use of the enormous information within the organization for clinical trials might be the first step, but it is not limited and may have a wider scope to expand beyond clinical trials, for example - some leading pharmaceuticals are creating proprietary data networks to gather, analyze, share, and respond to real-world outcomes and claims data. As a result, they are integrating with providers and payers. Users can choose what information they want, and what might not be useful in their field of work. They do not have to search different documents on various locations or platforms and can quickly navigate from one study to another in few seconds.

Conclusion

Based on the FAIR (Findable, Accessible, Interoperable, and Reusable) framework, if converting clinical data to valuable information is made available in a single place, it could be extremely powerful. Firstly, data should be easy to find, both by machines and humans, then the data should be accessible but also protected through proper authentication. Additionally, it needs to be integrated with other relevant data and requires interoperation with applications and workflows. Lastly, they should be reusable, for different applications. Most of the metadata and individual participant data have not been shared frequently with the wider scientific community apart from publications in peer-reviewed journals. But the fact is that these peer-reviewed journals have only a small subset of data collected, produced, and analyzed over time. Based on the committee on strategies for responsible sharing of clinical trial data, Institute of Medicine, National Academic Press (US), 20 April 2015, more and more clinical trial data at different stages has been planned to be shared among other investigators. Furthermore, through a legally mandated electronic-surveillance system, FDA is investing the evaluation of electronic health records through the Sentinel Initiative, which links and analyzes healthcare data from multiple sources5. As part of this system, the FDA has now secured access to data concerning more than 120 million patients nationwide. Pharmaceutical companies are uncertain to invest in big-data programs, and hesitant about increasing interactions with regulators. Taking baby steps to choose a specific area of interest to conquer, might be useful to obtain long-term benefits. Different organizations must come together and reinvest their interests in building this data hub. As we sort this out more, it would help in supporting pharmaceutical companies to provide more evidence to regulators. If data sharing is done in an efficient, standardized, and user-friendly manner, some of the challenges faced could be curbed, and information could be made available in one place. All this could lead to faster, cost-efficient, and safer drug discovery which is the need of the hour.

About the Author

Divya Kasargod
Consultant

Divya has 9+ years of experience in life sciences and healthcare areas. Her industrial experience extends across different domains including clinical trials documents, drug safety, healthcare out comes, Electronic Health records, patient services, Hospitals, EDI-837, Medicare, Medicaid, prescriptions, pharmaceutical R&D, Pharma regulatory submission and tracking. She has been enhancing customer experience by delivering compelling business value within pharmaceuticals, medical device and health insurance domain. She is a certified scrum master.

References

1. NIH clinical trials.gov https://clinicaltrials.gov/ct2/resources/trends
2. https://www.thehrdirector.com/features/hr-in-business/turning-data-into-insight/ Turning data into insight, 8 August 2016
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6742739/ Secondary Use of Laboratory Data: Potentialities and Limitations, 1 August 2019
4. https://www.centerwatch.com/articles/12702-new-mit-study-puts-clinical-research-success-rate-at-14-percent/ New MIT Study Puts Clinical Research Success Rate at 14Percent, 5 February 2018
5. https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d# How big data can revolutionize pharmaceutical R&D, 1 April 2013

For more information, contact askus@infosys.com

Infosys®
Navigate your next