

Clinical Data Aggregation



Author:

Ben McGraw
Strategic Marketing
PerkinElmer, Inc.
Waltham, MA

Background

With soaring clinical trial costs and complexity, the biopharmaceutical industry is constantly seeking new approaches to improve efficiency and productivity. The estimated cost of bringing a drug to market in the U.S. according to JAMA is \$1 billion¹, with easily more than half of the cost tied up in clinical development. At the same time, clinical trial sponsors and contract research organizations (CRO) are facing intense pressure to lower costs, speed the entire drug research and development process, and make clinical trials safer, more accurate, and efficient. However, with larger, longer, and more complex clinical trials, along with more stringent regulatory requirements and oversight, integrating and managing the vast amounts of data arriving at a central integration point from many different trial sources has become a daunting task.

The good news is that innovations in clinical trial technology are underway, promising to improve trial efficiency, accuracy, and the bottom line. Sponsors and contract research organizations (CRO) are seeking a more effective, comprehensive solution through a more-effective electronic data management strategy that includes an aggregated data repository solution with analytics and real-time access to the fully integrated data. Electronic data capture (EDC) and clinical trial management systems (CTMS), while preferable to paper-based systems, lack the full data integration capabilities necessary for effective R&D oversight - particularly for the conduct of clinical trials and the achievement of milestones.

One of the industry's key areas of focus for cost reduction is on-site monitoring, which is typically the most costly aspect of clinical trials, costing a whopping 25 to 50 percent of the clinical trial budget.² Through new approaches to monitoring that rely on advanced technology with data integration at the center of the solution, sponsors and their contract service providers are hoping to achieve greater efficiencies, while enhancing patient safety and improving the quality of clinical data.

Challenges

The intended goal of a fully integrated data aggregation solution is to continuously and automatically consolidate data from the variety of systems used to collect clinical, operational, safety, and outcomes data. Full aggregation of the vast amounts of data coming from many different sources and systems must take place before data analysis can begin. The ideal solution should automatically and continuously consolidate data from the many disparate systems to provide an enterprise-level analytics solution that empowers clinical development teams to make meaningful, data-driven decisions; to dramatically reduce the costs associated with clinical development; and guide new drugs and devices to market – faster.

Effective aggregation of so many disparate data sources to achieve this goal is extremely difficult. This is typically done manually or ends in failed attempts of developing custom data aggregation solutions.

From a source system perspective, many Sponsors and CROs have acquired a litany of on-premise or SaaS-based systems from vendors. Some Sponsors and CROs have implemented integrated software suites that have issues of product aggregation outside the product suite, since the products were not developed to work together. Sponsors and CROs also rely heavily on external data from vendors and partners. Often Sponsors and CROs have data silos created from historical systems or systems from acquired companies that require migration and manipulation of the data. A change in one source system requires time and investment to modify

and re-validate the data aggregation solution. Companies that perform data aggregation in-house often find that this results in data duplication, time intensive and costly migration, and new sources of risk.

Another problem when attempting to perform data aggregation is that all the various sources storing data separately may refer to the same item with a different name, presenting a major challenge for data aggregation as few sponsors and contract research organizations have master data management policies and technology in place.

The Typical Data Aggregation Solution

Many companies try to implement archaic data warehouses to address the clinical data aggregation challenge. Data warehouses leverage relational-databases built on hard-coded schemas that must be defined upfront. A data warehouse can take months or years to build because the schema must be defined up-front along with the questions that could be asked of the data. Unlike performing data analysis on manufacturing systems or various enterprise resource planning systems, drug development is constantly changing. Thus, the data and systems constantly change and in turn, the data warehouse schemas must be regularly modified and re-validated. This is not a flexible model with the rapidly evolving analytics needs.

What is the Ideal Clinical Data Aggregation Solution?

With the limitations on a relational database-oriented approach or semantic web-based technologies that consume unstructured data requiring semantic structuring on the back end, a point in the space between the two is needed. The ideal data aggregation platform is a cloud-based, horizontally scalable computing platform upon which analytic discoveries into the full research and development life cycle can be performed on the aggregated data.

This cloud-based data integration platform consumes data from many data sources leveraging a design that utilizes distinct technologies to handle each data source and type of data. By recognizing the different data requirements, the most appropriate informatics technologies are leveraged with an integrated polyglot database back-end.

Tailored to the needs of complex clinical data analyses and workflows, the core functionality is presented through an intuitive web interface to enable clinical development teams to perform tasks such as finding clinical subjects or samples of interest and decorating those subjects or samples with any number of clinical, operational, safety or outcomes data.

The assembled data can be returned into fit-for-purpose TIBCO Spotfire® templates designed to drive rapid insights into risk-based monitoring, critical operational metrics such as tracking key milestones across the entire portfolio of clinical candidates or improved interpretation of clinical data.

The clinical analytics solution is capable of analyzing data from multiple distinct clinical data sources while leaving the original source systems in place. Data is not transformed to a common data model, but left in the same data models found in the underlying transactional systems for a single source of truth.

The solution continuously and automatically consolidates data from the many disparate clinical development systems used to collect data related to clinical trials allowing for functions such as trial oversight through real-time risk assessment. It can analyze standardized datasets such as CDISC SDTM, CDISC ODM or flat files for analysis, is flexible enough to accommodate multiple disparate source systems and formats, and enables users to visualize outliers and trends.

Interactive data visualization and analysis solutions developed by PerkinElmer Informatics have helped companies streamline clinical, operational, safety and outcomes initiatives providing timely information and actionable insights that improve clinical trial management with solutions that can be tailored for medical, safety and operational management.

References

1. BIOPHARMADIVE. March 3, 2020. Accessed at: <https://www.biopharmadive.com/news/new-drug-cost-research-development-market-jama-study/573381/>
2. Ray S. Clinical Teams Should Re-Think Risk-Based Monitoring Costs to Improve Their Bottom Line. Accessed at: <http://www.cuttingedgeinfo.com/2013/risk-based-monitoring-costs/>