Ensuring Data Quality in a Complex Trial Landscape

Consolidated data review and artificial intelligence technologies are integral to managing high-volume, complex data generated from decentralized clinical trials.

Wendy Morahan
Sr. Director, Clinical Data Analytics
IQVIA Technologies

Jennifer Aimone
Director, Clinical Data Analytics
IQVIA Technologies

The medical research industry has been primed to move toward decentralized clinical trials (DCTs) in recent years since trial sites, investigators, and patients all have shown preference for DCTs in one form or another. The COVID-19 pandemic accelerated this movement, resulting in the shift toward a larger quantity of data source sites, greater data volume, and growing intricacy of collected data overall. At present, conventional data management architecture is far from suitable for ingestion, standardization, review, and identification of the large volume of data generated by DCTs. Clinical trial organizations can take steps to bring the scalability necessary to accommodate data handling requirements through approaches like consolidated single-point data review and incorporation of smart technologies to automate review and management. Such changes are critical for handling the emerging big data scale of the clinical trial environment through enhancing data management efficiency while improving data auditing, shedding redundancies, and applying technologies like artificial intelligence (AI) and machine learning (ML).

CLINICAL TRIAL CHANGES IMPACTING DATA MANAGEMENT

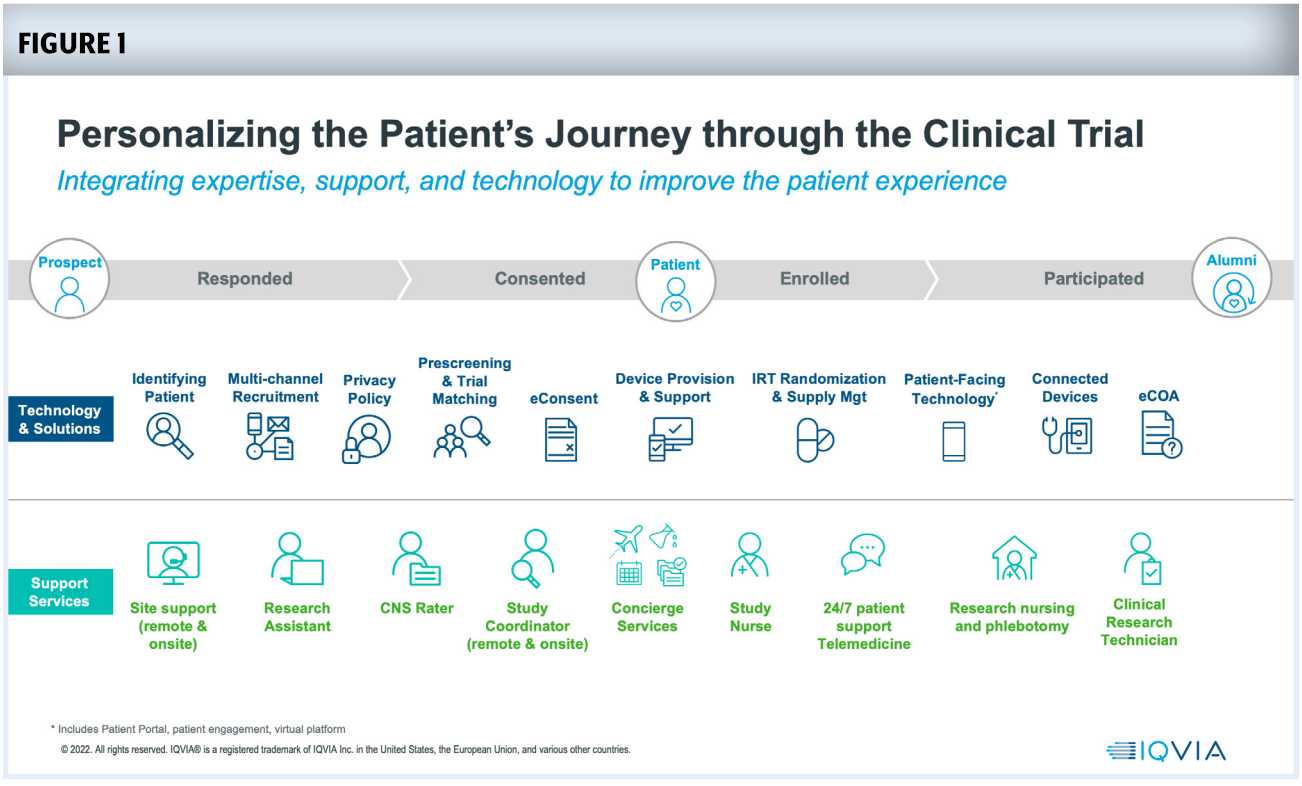
The clinical trial ecosystem trajectory for the near future is anticipated to present new industry challenges, including greater data complexity and volume. The evolving complexity and quantity of trial data can be traced to a confluence of several factors such as growing use of innovation technologies, biomarkers and 'omics analyses, novel study designs, connected devices, and decentralized trials, although the shift toward DCTs specifically has been a major driver of data intricacy and expansion.

While the shift toward using DCTs versus conventional trial formats was catalyzed by the pandemic, the favorability of the DCT format with both research organizations and patients alike is a component of continued DCT popularity.

Organizations have demonstrated interest in DCTs through investment choices such as the $1.9 billion in disclosed investments made in virtual clinical trial companies since 2020. Survey data from polling the IQVIA patient and trial site communities found 94% of investigators surveyed were in favor of either fully decentralized or hybrid DCT structures. Among investigators surveyed, more than 70% agree that the DCT approach relieves some degree of participation burdens on patients and thus can be more inclusive to people who would be unable to participate otherwise. While previous barriers to patients favoring DCTs have included unfamiliarity with technology like telemedicine and smart devices, IQVIA's patient surveys indicate that advances in consumer technologies like smartphones have made these concerns less likely to interfere with DCT patient participation. Among those surveyed, over 82% are comfortable using a mobile device and more than 80% are comfortable monitoring and self-reporting symptoms through an online application. Greater technology literacy among the patient populace has also brought greater expectation that clinical trial participation would incorporate technology-based flexibility. About 68% of IQVIA's surveyed patients expected some form of remote option like telehealth or in-home care would be available during clinical trials.

In mapping out a patient's journey through clinical trial participation, each stage of the experience will involve using specific technology and corresponding support services to ensure the step is completed properly as shown in **FIGURE 1**. Not only are multiple processes necessary at every level to make each stage operates in concert with the others, but each stage also represents a point in time where data is collected. Presently, research organizations manage data at these varying points of capture and use electronic data capture (EDC) to generate data lakes. This approach is used because data management capabilities are within the EDC systems themselves, so data cannot be managed independently of the EDC systems. While this approach can be workable for small datasets, the method is by no means scalable or sustainable as large datasets quickly become unwieldly.

FIGURE 1

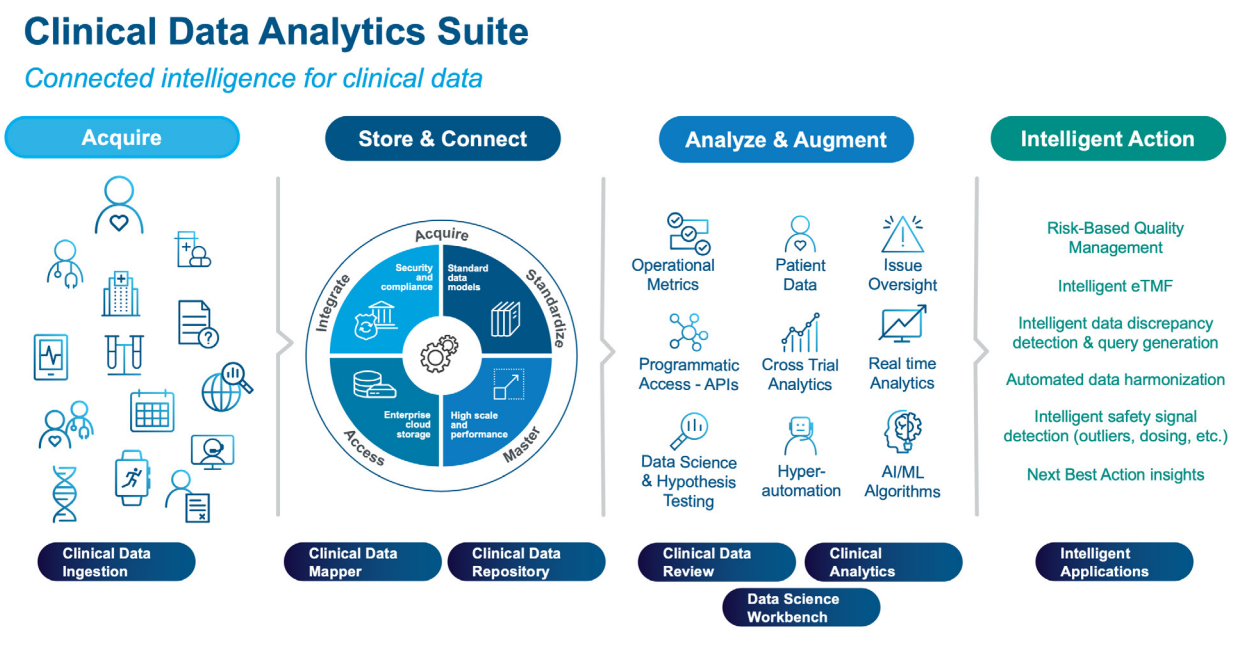


For the rapidly expanding data volumes collected in clinical trials, an entirely new data management paradigm is necessary. A streamlined solution would use a single platform for data acquisition that is fully agnostic to the data origination source while allowing for consolidated data management and review using smart automation tools. To address the need for such a unified data acquisition ecosystem, IQVIA developed the Clinical Data Analytics Suite that is diagrammed in **FIGURE 2**, with the general applications within clinical trial data management shown and the corresponding software titles below each column. The source-agnostic data collection entity within IQVIA's Clinical Data Analytics Suite is a cornerstone of the platform by allowing for consistent data ingestion gathered within a unified location and harmonized across data type for consolidated review. Keeping data within a central location then permits research team members to begin review, cleaning, issue management, and analytics with less dependency on IT for data integration near the end of the pipeline. Highly consistent data ingestion and CDISC-compliant mapping allow for more precise analyses of high-volume, complex datasets with reduced duplication through time-saving cross team information sharing. The unified ecosystem would also provide full transparency for auditing the workflow and yield nearly real-time access to data-driven business intelligence. Such features can be further optimized through AI and ML capabilities in the IQVIA platform to identify intelligent insights and actionable items from historical content for workflow improvement.

EFFICIENCY GAINS FROM DATA CONSOLIDATION FOR SINGLE-POINT REVIEW AND MANAGEMENT

The clinical data review process involves a rigorous assessment of data to affirm information veracity, identify discrepancies requiring correction, extract actionable insights, and monitor study progress. Additionally, the sensitive nature of human medical studies merits careful data assessment in clinical trials to ensure patient safety and monitor drug efficacy. Several different types of teams

FIGURE 2



can be necessary to perform clinical data review depending on the goals within the process. While some checks require decision-making from a human with medical training to identify data problems or safety issues, other portions of data review, such as steps that are involved in pulling data through the review workflow, can be handled with AI or ML using automatic checks at set points. As the data quantities and number of associated sources continue to grow, hiring additional people for data review becomes unsustainable and cost prohibitive. Although the complexity of reviews and requirement for human input can vary depending on the trial, there are still ways that advances in technology for big data platforms, analytics, predictive models, natural language processing, and ML can make major inroads into removing the burden from team members.

A conventional data review framework follows the process diagram in **FIGURE 3**, which shows a generalized clinical trial data review organizational structure and workflow. The far lefthand side of **FIGURE 3** represents the flow of data from *n* number of source systems to each group involved in data review. Each group contributing to data review receives a copy of the dataset and goes on to separately complete data ingestion, standardization, review, and insight identification before the multiple versions of processed data are reconciled and can be passed on to the customer. As a result, the workflow suffers from inefficiencies because the data are essentially decoupled from process by way of multiple data versions created and maintained throughout much of the review. This approach is prone to generate a high quantity of redundant reviews and queries, which causes both wasted time and great difficulty for auditing purposes. Moreover, as the number of source systems is scaled up, the data review model in **FIGURE 3** becomes untenable.

Given the inherent problems with attempting to scale up the traditional mode of data review, IQVIA redesigned the process by reconciling the new parameters set forth by DCTs, hybrid

FIGURE 3

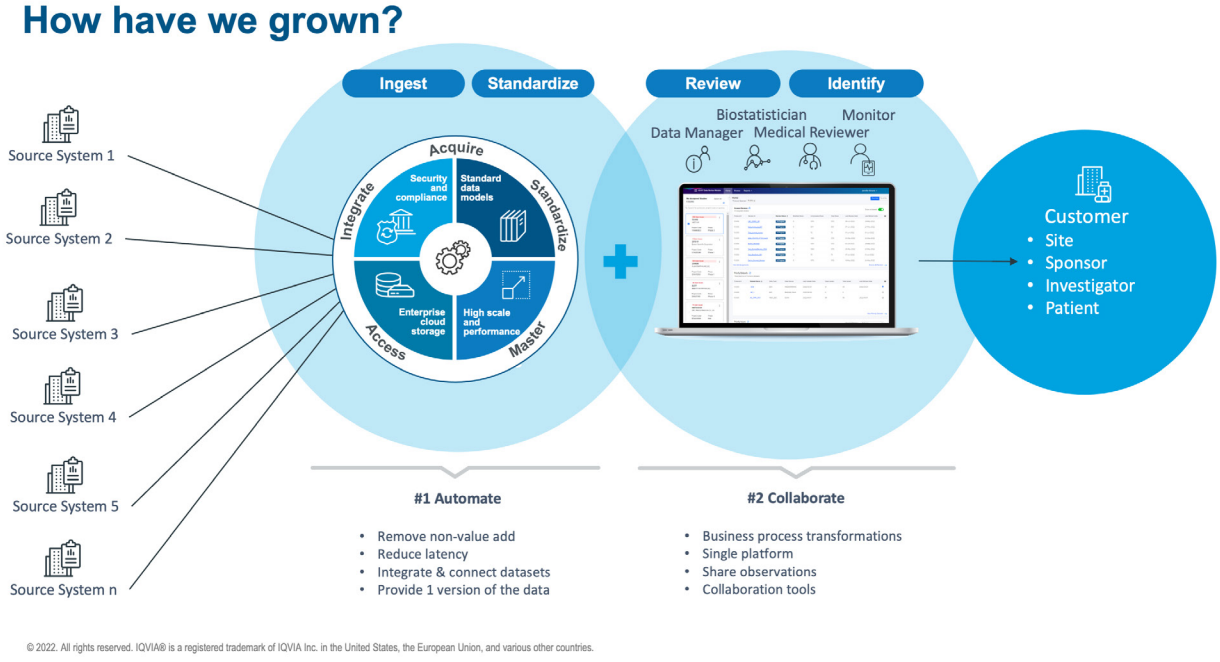
trial modes, and greater data complexity with a streamlined format incorporating key points for smart automation (**FIGURE 4**). To manage the continually increasing number of source systems, DCT data structures must be reinvented by integrating the movement of data with data processing. This occurs by first ensuring that all source systems report data to a centralized, automated system to ingest and standardize the data, represented on the lefthand side of **FIGURE 4**. Data passed into the centralized hub to be ingested and standardized are then sent to a single consolidated review platform that is accessible by data managers, biostatisticians, medical reviewers, and other review team members at once to perform data review and insight identification without duplicating queries, comments, and efforts via shared observation and collaboration tools (righthand side of **FIGURE 4**). With this structure, the review teams can focus efforts on outcomes like data reliability, accuracy, and scientific integrity rather than the non-value-added activities involving data pulls and technical activities that can be either taken on by someone else or automated.

SHAPING BETTER OUTCOMES THROUGH AI-POWERED DATA HARMONIZATION AND MANAGEMENT

In traditional data review, teams of reviewers are siloed into skill groups that comb through each piece of data followed by comparison of details, discrepancies, and insights on the back end. The redundant efforts involved in this approach can no longer suffice in a world where new forms of data capture and sources, digital health wearables, and connected devices heighten the complexity and volume of data.

Reconfiguring the review workflow to be centered around a single-point review will bring scalability without sacrificing data quality. Clinical trial data management must transition to being viewed as a large-scale operation dependent upon AI and ML intelligence for generating insights that empower data reviewers to focus on events and outcomes. With input contained and centralized, data can be mined through smart tools identifying insights and patterns, forecasting possible safety events, expanding scenarios for review, and predicting events not yet

FIGURE 4



reported. The accumulation of information libraries coupled with natural language generation can allow AI to automatically trigger queries and make recommendations based upon history. Moreover, smart tools can enhance the reviewer experience through learning team members' behavior to automate simple tasks and make recommendations based upon prior actions with continual refinement from reviewer feedback. The value and transformational impact of automation grows as clinical data management (CDM) evolves to clinical decision support (CDS) as conveyed in **FIGURE 5** regarding automation of traditional reviews, actional review, guided review, and supervised review. The pathway taken from CDM to CDS with smart technologies can vary for each organization and may depend upon the organization's data assets and technology maturity. AI can serve as a software tool with human-like behaviors, including self-optimization and learning, while robotic process automation (RPA) can perform basic automation of process-driven, predictable tasks using a decision tree. Intelligent process automation (IPA) can complete non-routine tasks combining components of RPA and AI to track information exchanged between entities in real time with intelligent, rules-driven decision making. Smart technologies such as AI, RPA, IPA, and ML are capable not only of pulling data through a CDM workflow in an automated fashion but also can identify meaningful observations and insights in a timely manner within ever-expanding clinical trial datasets.

CONCLUSION

Seismic changes to the clinical trial landscape have significantly impacted traditional data management practices to the point of compelling reassessment of the conventional data review ecosystem and workflow. Tightening the data review process into a consolidated, single point of review and management across all data sources is a key approach for creating the necessary increases in data handling efficiencies that maintains the high level of data quality expected in clinical trials. Additionally, the centralized review structure enables the use of AI and ML tools to ingest and harmonize data while empowering data review teams to shape better outcomes by optimizing the review process.

FIGURE 5