Paper LI04

Leveraging Data Science & Analytics to Optimize Clinical Trial Execution

Melanie Hullings, Formation Bio, New York, NY, USA
Laura Fazio, Formation Bio, New York, NY, USA

ABSTRACT

Through innovative leadership, we transformed a Clinical Data Management team by implementing open-source R programming for automated data cleaning and reporting. This evolved into two interconnected teams: Clinical Data Management and Clinical Data Analytics & Reporting, collaborating to efficiently execute clinical trials. Building upon this foundation, we are now establishing a pioneering Clinical Data Science program. By applying cutting-edge machine learning techniques, we aim to proactively address critical clinical development questions, enabling data-driven optimization of our clinical trial processes. This presentation will showcase our journey of innovation, discussing the strategic vision, practical implementation, and tangible impact of leveraging data science to streamline operations. We will share insights on fostering a culture of continuous improvement, upskilling team members, and harnessing advanced analytics to drive success. Attendees will gain valuable takeaways on leading transformative initiatives, empowering teams to push boundaries, and unlocking the full potential of data science in the clinical domain.

INTRODUCTION

In the rapidly evolving landscape of pharmaceutical research and development, the efficient execution of clinical trials remains a critical factor in bringing new therapies to patients. Traditional approaches to Clinical Data Management (CDM) have struggled to keep pace with the increasing complexity and volume of data generated in modern clinical trials [1]. This paper outlines our journey of transformation at Formation Bio, detailing how we leveraged data science and analytics to optimize clinical trial execution.

Our purpose is threefold:

1. to demonstrate the impact of implementing open-source R programming for automated data cleaning and reporting in CDM;
2. to showcase the evolution of our team structure and the establishment of a Clinical Data Analytics & Programming (CDAP) team;
3. and to provide insights into the strategic vision, practical implementation, and tangible outcomes of integrating advanced analytics into clinical trial processes.

HISTORICAL CONTEXT AND TRANSFORMATION

The evolution of CDM has been marked by several key milestones, from paper-based records in the 1920s to the introduction of Electronic Data Capture (EDC) systems in the late 1990s [2]. The digital transformation of CDM has been driven by improved data quality, increased efficiency, and the proliferation of advanced electronic systems [3]. Despite these advancements, many organizations still grapple with the challenges of managing and deriving insights from the vast amounts of data generated in clinical trials that require strict controls.

At Formation Bio, our transformation journey began with the strategic decision to implement open-source R programming for automated data cleaning and reporting. R has become increasingly popular in the pharmaceutical industry, with over 60% of data science job postings in the sector listing R as a required skill [4]. This choice was driven by R's extensive statistical capabilities, its extensibility through user-created packages, powerful data manipulation tools, support for reproducible analyses, sophisticated data visualization capabilities, and easy integration with other languages and tools commonly used in clinical research.

Our implementation process was methodical and comprehensive. We began with an initial assessment of our existing processes to identify areas where R could provide the most significant improvements. This was followed by a comprehensive team training program in R programming, focusing on packages most relevant to clinical data management. We then initiated pilot projects to test R-based solutions and demonstrate their potential. These successful pilots paved the way for full-scale implementation across our clinical data management processes. Throughout this journey, we established a culture of continuous improvement, constantly refining our R-based workflows based on feedback and emerging best practices.

The success of our R implementation catalyzed a broader transformation in our team structure. We evolved from a traditional Clinical Data Management team into two interconnected units: Clinical Data Management (CDM) and Clinical Data Analytics & Programming (CDAP). The CDM team focuses on ensuring data integrity, overseeing data collection, and maintaining regulatory compliance. Meanwhile, the CDAP team specializes in leveraging R and other analytical tools to provide advanced data analysis and dynamic reporting capabilities. This restructuring aligns with industry trends towards more specialized and analytics-focused roles in clinical research [5]

OVERCOMING TECHNOLOGICAL CHALLENGES

Our transformation journey was not without challenges. We encountered and addressed several technological hurdles that are common in the industry but often impede progress. For our Clinical Data Managers, key challenges included data integration and interoperability, maintaining data quality and consistency, effective use of EDC systems, ensuring regulatory compliance, and safeguarding data security.

To address these challenges, we utilized data integration platforms and CDISC standards to harmonize data across different systems. We implemented automated quality checks and regular audits to maintain data integrity. Comprehensive training programs and continuous support were provided to ensure effective use of EDC systems. We kept our team updated with the latest regulatory guidelines and used compliance-checking software to maintain adherence. Advanced encryption methods and secure access controls were implemented to protect sensitive clinical data.

For our Clinical Data Analysts, the challenges were centered around handling the increasing volume and complexity of data. We needed solutions for real-time data processing, big data analytics, effective data visualization, implementation of advanced statistical methods, and data standardization across studies.

To overcome these challenges, we adopted real-time processing frameworks and streamlined our data warehousing capabilities. We leveraged advanced analytics tools and machine learning algorithms to handle large, complex datasets. Sophisticated visualization tools were employed to create clear, impactful representations of data. We provided continuous education in the latest statistical techniques to our team and applied standardization techniques and common data models to ensure consistency across studies.

IMPACT AND RESULTS

The implementation of our new approach has yielded significant benefits across our clinical trial processes. We've saved the entire study team about 100 hours per week that were typically spent aggregating, analyzing, and reporting metrics that are now readily available in automated dashboards. We've seen a reduction in manual data cleaning time, allowing our CDM team to focus on more complex, value-added tasks such as issue resolution and prevention. The time to database lock has decreased by at least 50% since much less last-minute data reconciliations are needed, accelerating our ability to analyze and report trial results. Cross-functional team engagement in data review processes has improved, fostering better collaboration and more informed decision-making.

From a financial perspective, we've achieved an estimated 50% reduction in overall data management costs per study. This cost saving, coupled with the acceleration of trial timelines, has significantly improved our operational efficiency. Perhaps most importantly, we've seen a reduction in data queries and an improvement in overall data quality scores, enhancing the integrity and reliability of our trial data.

As a concrete example of our automation efforts, we developed an R program to detect protocol deviations. This program analyzes incoming data in real-time, flagging potential deviations for review. It doubled the deviation detection rate while reducing the time spent on manual reviews by 55% on the first study where it was implemented. In subsequent studies, automated deviation detection rates are over 80%. This not only improves the quality of our trials but also enhances patient safety by allowing faster responses to potential issues.

PATH FORWARD

Building on our success with R programming and advanced analytics, we are now poised to significantly expand our data science capabilities. We are establishing a comprehensive Data Science program that incorporates advanced machine learning techniques, predictive modeling, and artificial intelligence to proactively address critical clinical development questions.

Key components of this program include the implementation of an advanced analytics infrastructure using scalable cloud-based solutions and data lakes. We are developing machine learning (ML) models for patient recruitment optimization, automated anomaly detection in clinical data, and natural language (NLP) and artificial intelligence (AI) processing for efficient handling of unstructured data from medical records and adverse event reports. Real-time

analytics and visualization capabilities are being enhanced to provide immediate insights into trial progress, patient safety, and data quality.

Our AI integration strategy follows a phased approach, beginning with process automation of routine tasks, followed by the implementation of intelligent assistants to support daily activities of our clinical data teams. The later phases involve developing predictive analytics models to forecast potential issues in trial conduct, and ultimately, creating semi-autonomous systems capable of managing certain aspects of clinical trials with minimal human intervention, always under strict regulatory compliance and human oversight. As we expand our data science capabilities, we remain steadfastly committed to maintaining the highest standards of data privacy, security, and regulatory compliance.

To support this ambitious expansion, we are investing heavily in training and talent development. This includes advanced training programs in data science, ML, and AI for our existing staff, recruitment of specialized data scientists and collaborations with academic institutions to stay at the forefront of clinical data science innovations.

CONCLUSION

Our journey at Formation Bio demonstrates the transformative power of leveraging data science and analytics in clinical trial execution. By implementing open-source programming, evolving our team structure, and embracing advanced analytics, we have significantly improved the efficiency and quality of our clinical data management processes.

As we look to the future, we see AI and advanced data science as the next frontier in optimizing clinical trials. We envision a future where all roles in biometrics are supported by ML and AI, leading to more efficient processes, immediate issue triaging, proactive study operations, and continuous learning.

Our experience underscores the importance of fostering a culture of continuous improvement, investing in team member skills, and maintaining a balance between innovation and regulatory compliance. As the field of clinical research continues to evolve, we believe that embracing data science and analytics will be crucial for organizations seeking to stay at the forefront of drug development and bring life-changing therapies to patients more quickly and efficiently.

REFERENCE

[1] Kush, R. D., Warzel, D., Kush, M. A., Sherman, A., Navarro, E. A., Fitzmartin, R., Pétavy, F., Galvez, J., Becnel, L. B., Zhou, F. L., Harmon, N., Jauregui, B., Jackson, T., & Hudson, L. (2020). FAIR data sharing: The roles of common data elements and harmonization. Journal of biomedical informatics, 107, 103421. https://doi.org/10.1016/j.jbi.2020.103421

[2] Krishnankutty, B., Bellary, S., Kumar, N. B., & Moodahadu, L. S. (2012). Data management in clinical research: An overview. Indian journal of pharmacology, 44(2), 168–172. https://doi.org/10.4103/0253-7613.93842

[3] Richesson, R. L., & Andrews, J. E. (Eds.). (2019). Clinical research informatics (2nd ed.). Springer Nature.

[4] Glassdoor Economic Research. (2023). Data science in pharma: 2023 report. Glassdoor.

[5] Society for Clinical Data Management. (2023). Annual survey on clinical data management practices. SCDM

ACKNOWLEDGMENTS

The authors would like to acknowledge the large cross-functional effort across teams at Formation Bio. as well as with vendors, to implement the systems and tools required to utilize open-source programming methods in a Good Clinical Practice (GCP) context.

RECOMMENDED READING

R: Regulatory Compliance and Validation Issues A Guidance Document for the Use of R in Regulated Clinical Trial Environments. (https://www.r-project.org/doc/R-FDA.pdf)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Melanie Hullings
Company: Formation Bio
Email: melanie@formation.bio

Author Name: Laura Fazio
Company: Formation Bio
Email: lfazio@formation.bio