PharmaSUG 2024 - Paper LS- 443

Data Harmony Revolution: Rocking Trials with Clinical Data Literacy

Melanie Hullings, Emily Murphy, Andrew Burd, and Helena Belloff, Formation Bio

Abstract

Clinical Data Analytics and Programming (CDAP) is centered around the collection, monitoring, reporting, and analysis of clinical trial data. Our team's objective is to improve the efficiency of clinical trials by producing high quality, reliable, and statistically robust clinical datasets. In order to accomplish this, analysts on the team must have an understanding of all aspects of technical frameworks and clinical trial operations, including medical and safety monitoring. At the same time, clinical trials require a large, multi-disciplinary team across many different highly-specialized clinical, operational, and technical roles. Because of the varied backgrounds, it is hard for everyone to understand what the other teams are doing, interdependencies, and downstream effects of decisions.

In an effort to bridge any gaps in understanding across the interdisciplinary teams, CDAP ran a 12-week internal internship program open to all employees. The purpose of this program was to teach the regulatory and operational basics of clinical trial data collection and analysis, empower people to work with advanced technical tools, and develop programming skills they could use to augment their own workflows. The result of this program, which culminated in a final project presentation showcasing R programming and data analysis skills, was a workforce with improved clinical and coding skills, enhanced transparency and empathy between teams, and a foundation for internal and external training programs.

Lost In Translation

Successful implementation of technology in healthcare has long been recognized as a challenge, particularly because it requires extensive alignment between many different cross-functional stakeholders with very different expertise and priorities (Cresswell 2013). The challenges are most commonly seen in health systems where technology must be implemented effectively to manage patient care, as well as all the other parties involved across clinical, leadership, operations, and information technology departments. These same challenges are also experienced by the diverse interdisciplinary teams running a clinical trial in the pharmaceutical industry where medical, safety, clinical, and biometrics need to translate requirements across the entire clinical trial process into solutions for technical teams in order to effectively run the trial and collect quality data to support study endpoints. The success of these systems is entirely dependent on effective implementation and management to support workflows, though often these requirements get lost in translation (van Gemert-Pijnen 2022).

Biometrics is the Interpreter between Clinical and Technical Teams

Biometrics teams are charged with applying statistics and mathematics to biological and medical data. In order to be successful, members of biometrics teams need to have a deep understanding of the clinical trial process through industry experience and advanced degrees. On top of this, they also require technical skills and usually write code in one or multiple languages. Biometrics inherently has knowledge of the full lifecycle of clinical data from creation to interpretation due to their deep expertise.

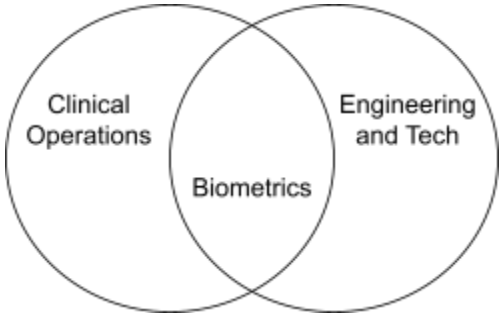

Figure 1. Skill Overlap of a Pharmaceutical Company Workforce

Comprising four essential functions - biostatistics, statistical brogramming, clinical data management, and clinical data analytics and programming (CDAP) - the biometrics team plays a multifaceted role in ensuring the integrity, quality, and analysis of clinical trial data. This can be a significant burden as they are ultimately responsible for managing and analyzing the trial results that clinical development relies on.

Biostatistics

At the core of the biometrics team lies the biostatistics function, a discipline vital for interpreting and analyzing clinical trial data with statistical rigor. Biostatisticians within the team are tasked with designing study protocols, determining sample sizes, selecting appropriate statistical methodologies, and analyzing study results. Their expertise ensures that clinical trial data is analyzed accurately and in accordance with regulatory standards, enabling robust and reliable conclusions.

Statistical Programming

Working hand-in-hand with biostatisticians, the statistical programming function translates statistical analyses into validated, executable code. Statistical programmers develop and implement data manipulation and analysis programs using specialized statistical software (SAS® or R), ensuring the reproducibility and efficiency of data analyses. Their role is critical in transforming raw clinical trial data into interpretable insights. Tangibly, the outputs of a Statistical Programming team are Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets and their accompanying Tables, Listings, and Figures (TLFs) that are used in regulatory submissions.

Clinical Data Management

The clinical data management (CDM) function is responsible for collection, storage, and quality control of clinical trial data. They oversee the design and implementation of electronic data capture systems, develop data management plans, and perform data cleaning, review, querying, and reconciliation activities. By maintaining high-quality and reliable clinical datasets, they ensure the integrity and accuracy of trial data, laying the groundwork for meaningful analyses and interpretation.

Clinical Data Analytics & Programming

The clinical data analytics & programming (CDAP) team was purposefully built at Formation Bio to primarily support the rest of the biometrics team, as well as other study team members, to automate data cleaning, reporting, and analysis to drive trial efficiency and data quality. CDAP also partners with engineering to build the clinical data platform and a continually growing set of data modeling and analysis tools connected to the validated end-to-end clinical data programming environment. This platform facilitates automated reporting, interactive data visualization applications, and clinical data science applications to support clinical development.

Clinical data science is an emerging capability within the CDAP team focused on leveraging data science methods including advanced analytics and machine learning techniques to derive insights from complex clinical data. This function employs statistical, predictive modeling, data mining, and artificial intelligence algorithms to identify patterns, trends, and correlations within large datasets. This role is critical in transforming raw clinical trial data into actionable insights and facilitating data-driven decision-making throughout the trial lifecycle.

The CDAP team is the most cross-functional of any team within the company and are required to translate across many levels of clinical and technical vocabulary to connect requirements to implementation and utility. As a result, other teams within the company are only exposed to a limited scope of the CDAP team’s capabilities and subject matter expertise. CDAP sought to cross-educate others across stakeholder teams and increase internal visibility of the scope of the team’s work in order to improve communication between teams and efficiency of the team’s own work.

The Internship

To bridge gaps in understanding across the interdisciplinary teams, CDAP ran a 12-week internal internship program open to all employees. The purpose of the program was to teach biometrics principles

to clinical and technical counterparts. The colleagues with more technical backgrounds could take away the regulatory and operational basics of clinical trial data collection and analysis. The clinical colleagues could be empowered to work with advanced technical tools and develop programming skills they could use to augment their own workflows.

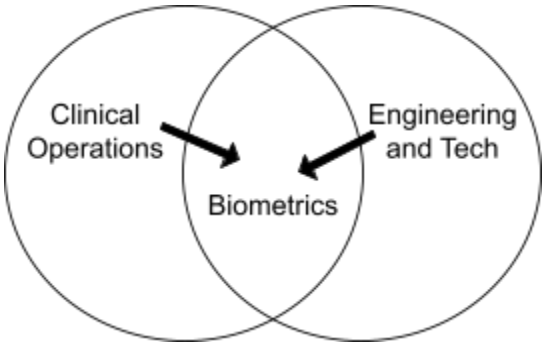

Figure 2. Goal of the Internship is to Cross-Train Technical and Clinical Teams

From Creation to Insights - Gaining an Understanding of Clinical Data Lifecycle

In order to demonstrate the full analysis lifecycle of clinical data for the biometrics team, the CDAP team created a pilot protocol and sandbox study in our electronic data capture (EDC) system to be able to walk through the end to end process that is typical for the biometrics team. Additionally, participants in the internship could walk in the shoes of both a clinical research coordinator (CRC) and a subject by being able to experience entering clinical data.

Figure 3. End to End Lifecycle of Biometrics workflow

Reading and Interpreting a Protocol

For those without a clinical background, a protocol can be an intimidating document. Those with experience running trials know the protocol is one of the most valuable resources to reference throughout the course of a trial. In order to bridge the gap in understanding, the internship provided a foundation of clinical trial basics covering topics such as the phases of a clinical trial and what primary and secondary endpoints are. Building on that foundation, one of the internship lectures was dedicated to strategies on how to read and interpret a protocol including the schedule of activities (SoA) and the inclusion/exclusion criteria.

Collecting Data in an EDC

Many members of a typical Biometrics team may even be unfamiliar with data capture workflows in an EDC system. Many downstream users of data do not have any insight into what happens at the data capture step in the process. However, hands-on exposure to data capture in an EDC system has many benefits for the downstream users of the data including creating an intuitive understanding of the structure of the raw clinical data.

Understanding Structure of Clinical Data

Data coming out of an EDC tends to be at a form level which can lead to an overwhelming amount of raw clinical data to sort through to those without subject matter expertise. Making quick sense of a complicated data model can be difficult. On top of that, SDTM and ADaM data models can have steep learning curves due to the number of rules and standards governing these data types. This phase of the

internship was very hands-on and relied heavily on the foundations of protocol interpretation and data collection to drive understanding.

Getting Started in R

R is growing in popularity as a tool used for clinical analytics and reporting. One benefit of leveraging R is the open-source community built around the tool. One side effect is that it can be hard to know where to start as a new programmer. There's a lot of foundation that needs to be built before these open source tools can be leveraged to their greatest potential.

Luckily there are more and more free courses available for a new clinical data scientist to be able to use to build up their R skills. Some favorites of the team include Swirl, Edx, and Coursera. We also created R scripts with exercises that walked participants through the basics of programming. Those ramping up their new R skills through the internship found success in implementing AI tools to help accelerate the process as well. ChatGPT has been hugely helpful in debugging code, translating code between languages, and helping to learn and implement new skills faster.

Harmonization of Data and Teams

Overall, the internship had ~25% of the company participating in the program with at least one representative from each department within the company including Engineering, Operations, Legal, People, Quality Assurance (QA) and Leadership. At the conclusion of the internship, participants spent the last few weeks working in smaller groups on projects using R to analyze the data generated by the test study in the EDC and presenting their data visualization and analysis products.

By upskilling study team members with less expertise to be able to perform some of their own ad hoc reporting and programming, this has also moved some of the simpler programming tasks to other teams so that CDAP can focus advanced technical data science resources on higher impact work. This has not only empowered less technical teams to more quickly discover insights on their own, but optimized the use of the CDAP teams expertise in advanced analytics to focus on higher impact data science projects.

Conclusions

The company-wide participation in this internal internship program demonstrated the strong interest in learning more about clinical data topics and tools across teams and successfully increased cross-functional visibility and understanding for our CDAP team. The continued cross-functional collaboration highlights how effective the internship was at promoting communication among teams.

Feedback collected through surveys and retros highlighted that the broad spectrum of topics we sought to cover across clinical data governance, capture systems, and analysis could be grouped together for clarity and specific use cases within our company. We have since separated clinical data topics into a higher-level onboarding session that covers regulations and systems and have separately implemented an R Programming workshops series and working group for all R programmers to collaborate weekly in working sessions.

Ultimately we achieved our goal of empowering clinical and technical teams through a shared understanding of both areas of expertise. Clinical teams are able to identify and help to implement more efficient, data-driven tools into their workflows and technical teams are more accurately able to translate requirements into better products and anticipate the needs of the clinical teams. We hope to continue to build a culture of data-driven decision making by continuing to find opportunities to bring together clinical and technical teams to bring new treatments to patients faster and more efficiently.

References

Cresswell, K. M., Bates, D. W., & Sheikh, A. 2013. "Ten key considerations for the successful implementation and adoption of large-scale health information technology." *Journal of the American Medical Informatics Association : JAMIA*, 20(e1), e9–e13. https://doi.org/10.1136/amiajnl-2013-001684

van Gemert-Pijnen J. L. 2022. "Implementation of health technology: Directions for research and practice." *Frontiers in digital health*, 4, 1030194. https://doi.org/10.3389/fdgth.2022.1030194

Recommended Reading

- *Swirl: https://swirlstats.com/*
- *Coursera: https://www.coursera.org/*
- *Edx: https://www.edx.org/*

Contact Information <heading 1>

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

Melanie Hullings
Formation Bio
melanie@formation.bio
https://www.formation.bio/

Emily Murphy
Formation Bio
emurphy@formation.bio
https://www.formation.bio/

Andrew Burd
Formation Bio
aburd@formation.bio
https://www.formation.bio/

Helena Belloff
Formation Bio
helena@formation.bio
https://www.formation.bio/