Driving Clinical Data Value with Data Science

To capitalise on the digital boom within pharma, companies need to steer their focus towards forming key insights from quality research data, with skilled data teams at the wheel

Richard Young at Veeva Systems

When you have disparate clinical data sources, it can be difficult to spot patterns and reconcile differences. Mapping multiple sources of data that are in different formats is time-intensive and introduces traceability challenges. It requires the application of data science to gather, harmonise, and properly format this information for research scientists so they can analyse clinical trials effectively and present their results. For organisations to view and analyse big data in this way, however, they will need to deploy the appropriate technology and train their data managers.

Excellence in Data Science

When connectivity across data sources in clinical trials is lacking, it creates quality issues that can slip through data management, impacting downstream users such as medical monitors. These issues are costly because the decisions made by medical monitors using poor quality data can lead to false positives or false negatives, potentially costing a pharmaceutical company millions of dollars (1).

There is transformation happening in data management as companies shift from a manual process to modern systems. The industry is leveraging machine learning (ML) to automate the mapping of data into the study data tabulation model (SDTM) format and robotic process automation (RPA), which can help code incoming data.

This is providing data managers with the tools needed to clean and harmonise data in the fastest and most cost-effective way. By adopting this data science approach, data managers can automate manual tasks, streamline processes, and present their organisation with the most relevant insights.

There is also a desire to have real-time visibility into patient data, to help organisations run more effective and efficient clinical trials. First, it helps to identify compliant patients and decrease loss to follow-up, reducing the number of patients needed and speeding database lock. Secondly, enrolment dashboards will show site performance, helping to highlight what is or is not working earlier in the process so other sites can correct it. Finally, data visualisations of patient data, such as blood pressure readings, can reveal outliers or potential adverse events.

Clean, reliable data are a prerequisite to realising these advantages. Getting that data to decision makers faster will allow them to proactively manage trials, accelerate execution, and increase efficiency.

Emerging technologies, such as artificial intelligence (AI), ML, sensors, and wearables, could also help to design more effective trials and improve patient safety, but they are yet to live up to their promise. A key reason for that lack of success using them is that the heterogeneity and discord in clinical data poses a challenge to AI or ML projects. If organisations don't find an effective way to manage and harmonise data from disparate sources, this potential will be hard to realise.

Real transformation occurs when data managers can apply data science to master the idiosyncrasies of biologic and real-world data. The ability to clean data will provide confidence in patterns and outcome predictions, driving better results and speed in trials.

Arming Data Scientists With Technology

Traditionally, the primary source and store of clinical data has been case report forms (CRFs) in electronic data capture (EDC) systems. In most organisations, the EDC remains the sole system designed to manage clinical data. Yet, these sources can no longer be the centre of a data manager's universe, with industry leaders estimating that less than 30% of data volume now comes from the EDC. A separate 2017 report by Tufts Center for the Study of Drug Development also found that most organisations (77%) experience difficulty loading external data sources into their EDC (2). Two-thirds (66%) attributed

those difficulties to EDC system limitations and integration issues.

Organisations looking to invest in their clinical data infrastructure should therefore explore technologies that offer:

- **A platform for clinical data management:** A purpose-built platform to aggregate, clean, and normalise data will help data managers work more efficiently. A dedicated platform for managing all data sources was previously seen in the industry as a luxury, but this should now be considered fundamental. Two of the essential steps for data management are selecting the right data aggregation platform and the learning required to use the query and reporting capabilities. The capabilities will automate or eliminate manual data cleaning tasks. Advanced reporting will also equip data managers to deliver high-value clinical data insights, and free up their time

- **Scripting and visualisation:** Visualisation capabilities help you look at data and extract meaning from the information. This level of usability is key for data managers as it enables them to move quickly and explore different aspects of a problem. In addition, visualisation makes it easier to explore other ideas and identify the right questions to ask

- **Augmented intelligence:** Applications such as ML, natural language processing, and robotic process automation (RPA) are crucial tools for a data scientist. Their deployment should be viewed as a way to help the data management team (not as a means to replace that team). Data scientists play an essential role in training the learning algorithms and in enabling these systems to deliver value

Skills for Success in Data Science

So, what does this mean for data managers today? We recommend four steps be taken to help data managers contribute as trusted advisors and help their organisation leverage its clinical data:

1. **Embrace real-world data:** Research the specific ALCOA (attributable, legible, contemporaneous, original, accurate) risks and limitations associated with each real-world data source. Patient registries, for example, provide extensive observational data that are of high quality and relatively inexpensive. However, confounding (i.e., the inability to isolate the impact of relationships between dependent and independent variables) is a significant problem and makes it difficult to attribute trends in registry data to particular therapies

2. **Get familiar with regulatory guidance:** Read the guidance on using electronic health record (EHR) data in clinical investigations, including recommended practices for everyday situations such as handling data modifications. EHR-EDC integrations reduce the need for data entry by sites and associated source data verification (SDV), but numerous complications exist. Data managers should advise the organisation on integration decisions that help preserve data lineage, protect personal identifiable data, and preserve identity masking

3. **Advise study teams on data source selections:** Identifying the appropriate data source requires making informed decisions and balancing trade-offs, such as the sources, which contain the desired information with the greatest reliability and fewest gaps, in high enough volume to balance across treatment arms. Educational courses from *The Journal of the Society for Clinical Data Management* (JSCDM), such as mastering mobile and digital technologies, provide a framework for evaluating new sources in clinical trials

4. **Take an active role in data management technology selections:** Many established systems were architected before the recent expansion in data sources. Data managers should research the capabilities of modern tools and platforms to identify those designed with diverse, heterogeneous data sources in mind. Industry conferences, such as the DIA and JSCDM annual meetings, are most valuable during times of change. It is worth attending an industry conference to speak with early adopters and supporting technology providers

Challenges with data quality was the third most-cited barrier to completing clinical trials, according to a study carried out by the PwC Health Research Institute, and data management is becoming more difficult with each new source of data added (3). And yet, innovation in the tools and training to support data managers has been lacking for too long. Biopharmaceutical companies can capitalise on the benefits of AI and real-world data if they develop a strategy that embraces data science and invest in the skills and technology needed to support this approach.

References

1. *Burt T* et al, *The burden of the "false-negatives" in clinical development: analyses of current and alternative scenarios and corrective measures,* Clinical and translational science *10(6), pp470-9, 2017*
2. *Visit: www.veeva.com/wp-content/uploads/2018/03/Tufts-Impact-Report-2017-eClinical-Landscape-Study.pdf*
3. *Visit: www.pwc.com/gx/en/healthcare/pdf/global-top-health-industry-issues-2018-pwc.pdf*

Richard Young is Vice President of Clinical Data Strategy at **Veeva Systems**. With almost 25 years of experience in life science, Richard is known for his executive vision and proven operational experience in data management, eClinical solutions, and advanced clinical strategies.

richard.young@veeva.com