

Strengthening eClinical Data Integrity: Data Monitoring

By Terek Peterson, MBA, Vice President Customer Experience and Data Science

As the use of eCOA expands to incorporate more submission-critical elements, especially key endpoints and factors contributing to per protocol analyses, customized compliance strategies are more important than ever to ensure data integrity. By following ALCOA principles for data integrity, quality assurance processes, system controls and customized approaches can drive completeness, consistency, and accuracy of data throughout the data lifecycle of a clinical trial.

System-level processes and custom functionality can help study teams manage risk of non-compliance through edit checks, data monitoring and predictive modeling.

Data Monitoring

Across clinical development, data monitoring is an expanding practice that applies targeted algorithms to enable rapid, automated data review that flags trends of interest or anomalies. It may also involve a combination of automation and clinical expertise for targeted, manual data review of select data points. Both techniques provide an early warning system before potential risks escalate into liabilities.

Within eCOA data sets, data monitoring can help identify compliance patterns that are below or near threshold levels, allowing for additional analysis by qualified staff to evaluate the scope and severity of the issue. Clinical expertise is needed to assess the issue in the context of protocol requirements. A clinical expert can determine if the issue requires a formal data correction or if it can be addressed in downstream statistical analysis.

Data monitoring will detect a range of findings, from simple anomalies to serious problematic issues, such as the otherwise compliant patient who starts a questionnaire on a long-haul flight, loses connectivity, and completes it upon landing, the next day. The date of record will be flagged as missing on one day or duplicate on another day. And then there are more complex issues, such as unblinded support for data and safety monitoring boards (DSMBs). What is more, ostensibly insignificant issues can add up to a big impact by the end of a trial.

When do anomalies become a trend, and how often should routine data review take place? These and other questions are best addressed upfront in the data management plan. Weekly reviews are standard practice, but biweekly reviews may be needed for pivotal studies.

Early identification also allows for intervention and corrective actions, to prevent problematic trends from escalating into larger challenges. For example, if a protocol requires 80% compliance and data monitoring shows compliance has been varying between 79% and 85%, a risk mitigation plan may be executed at that point to evaluate the likelihood of meeting the final compliance target, and to identify reasons for low compliance (such as demographic-related factors, problems at study sites or problematic issues with the outcome instrument).

Once enough data are gathered, predictive modeling can anticipate non-compliance situations and trigger preemptive actions. "Predictive modeling can provide an estimate of the likelihood of missing targets, helping to prioritize risk mitigation across multiple objectives," explains Dennis Sweitzer, PhD, a career biostatistician. For example, modeling and simulation of patient-level study events such as dropouts, SAEs, or study anomalies can estimate event frequencies and their likely impact early in the study, while corrective action is most productive.

The Future

Insights gained from data monitoring go beyond compliance to inform a range of other issues related to safety, data quality and operational performance.

Adverse Events

Various forms of data monitoring have long been used for signal detection and epidemiologic techniques to evaluate side effects of investigational and approved drugs. One emerging trend is the pairing of machine learning (ML) and natural language processing (NLP) with automated monitoring. Machine learning incorporates levels of supervision, with varying amounts of reference data to help determine patterns, relationships, and structures, typically on more granular levels than humans are capable of. NLP techniques are used to help answer specific problems and often involve text-mining of structured or semi-structured data. The database of research information is run through an NLP application with focused topics, specific phrases, or words for it to search for. Text-mining not only identifies facts, it analyzes to determine contextual information. For example, free text fields in data capture often have no relationship with entry inputs into other data capture systems. Potentially valuable information requires significant rework to retrieve data, analyze it and connect it to related systems, wasting research time and money. Using NLP, free-text fields can be mined for potential adverse events. Automated data extraction and codification can drastically reduce the typical time required for assembly and in-depth analysis of multiple data sources.

Continuous Improvement

Monitoring and simulation not only allow study teams to address current compliance or safety risks in real-time but also provide the ability to reduce future risks, and help studies run more efficiently through identification of outliers, such as missing data, consistency of frequently collected data, query rates and site-level or patient-level protocol deviations. Higher sensitivity analytical capabilities can also identify patterns that indicate near-failure. Insights gained from data monitoring in real-world usage can lead to better protocol and software design.

While applications and associated benefits of data monitoring and predictive analytics have long been used in other types of data collection, they are newer to eCOA. However, as big data takes shape and the industry further adopts mobile technologies and artificial intelligence on an enterprise scale, data integrity will be a key benefactor as these techniques are continually refined and put to everyday use.