Paper AR04

The Right Tool for the Job: Enhancing Risk-Based Monitoring and Data Cleaning

Casey Higgins, Atorus Research, Apex, USA
Pavel Burmenko, Veeva, Seattle, USA

ABSTRACT

Integrating advanced risk management and data-cleaning techniques is crucial for clinical trials. Visual analytics, like R Shiny, enable real-time interaction with data, which is ideal for initial exploration and monitoring key metrics such as adverse event rates. Visualizations highlight areas of concern quickly, aiding prompt decision-making.

Using a range of programming and statistical techniques, from basic to more advanced, data science tools can detect subtle anomalies not visible through manual inspection, automating the identification of outliers and inconsistencies. It's crucial to combine aggregate metrics with detailed reports to provide a granular view, supporting rigorous inspection, validation, and issue tracking. This thorough process ensures that no detail is overlooked, providing reassurance in the validation process.

Combining visual analytics with statistical/clinical programming creates a robust risk management strategy. This dual approach enhances data quality, streamlines the cleaning process, and reduces time to database lock, improving operational efficiency and decision-making.

INTRODUCTION

Clinical trials generate vast amounts of data from multiple sources, requiring robust tools and workflows to ensure data quality and regulatory compliance. The traditional approach of manually reviewing all data is no longer feasible. Instead, risk-based approaches combined with advanced tools like data workbenches and analytics platforms are transforming data management practices.

This paper explores the complementary roles of data workbenches and analytics tools in clinical trial data management. We discuss how these tools integrate with various systems, the importance of selecting the right tool for the task, and how modern Application Programming Interfaces (APIs) and GxP-compliant workflows improve efficiency and quality.

DATA WORKBENCHES: THE FOUNDATION OF DATA QUALITY

Data workbenches are ideally suited for aggregating and reconciling data from multiple source systems, including:

- Electronic Data Capture (EDC): Clinical data entered directly by investigator sites.
- Third-Party Providers: Central lab results, electronic clinical outcomes assessments (eCOA), and wearable device data.
- Clinical Trial Management Systems (CTMS): Operational data tracking trial progress, protocol deviations, etc.

These tools automate data reconciliation processes, which previously had to be performed manually or using offline programmatic tools (e.g. to map and match patient IDs following different naming conventions between sources and to identify value-level discrepancies). Value-level cross-source data checks are mandatory so that data flowing downstream is accurate and consistent. Historically, these cross-source checks were programmed manually in SAS on static data extracts, refreshed monthly by programmers and sent as spreadsheets to data managers and other stakeholders for offline review. This led to constant investigation and re-work due to one data source snapshot being out of sync with another. Modern data workbenches automatically refresh data checks whenever new data is loaded and can automatically send actions to the source system via API. For example, a data workbench can use APIs to open a data discrepancy query on EDC or third party data or to create a potential protocol deviation in a Clinical Trial Management System (CTMS).

A key advantage of data workbenches is their ability to focus on critical variables that directly impact trial outcomes or patient safety. By specifying risk-based approaches in the Data Management Plan (DMP), teams can avoid spending excessive effort on non-critical data. For example, an efficacy Time-to-Event analysis in an oncology study may require the censoring of certain events that are collected from EDC Case Report Forms (CRFs), electronic Patient Reported Outcomes (ePRO), and central laboratory data. The accuracy of these censoring algorithms are highly sensitive to any discrepancies related to patient identifiers, assessment date values or formats, and result values. Programmatic checks of this data are much more effective than visual review, especially when the checks can run automatically and trigger actions in source systems, such as queries and protocol deviations. This upstream cleaning increases the efficiency of analytics tools by delivering cleaner datasets, reducing the need for manual intervention during analysis.

ANALYTICS TOOLS: PURPOSE-BUILT FOR INSIGHTS

While data workbenches excel at presenting data in granular detail as listings for ensuring data quality, analytics tools excel at presenting data in the form of trend analyses utilizing advanced statistical methods and patient profiles that include reviewer-friendly graphical capabilities. These tools are not intended for data cleaning, but rather for deriving insights.

TREND ANALYSES

Trend analyses utilize advanced statistical techniques to calculate the rates of clinical and operational outcomes. For example, an RBQM Dashboard produced with R-Shiny can highlight statistically significant differences between investigator sites in a multi-center clinical trial. Review of data listings in a data workbench would not allow a reviewer to identify that one site's adverse event reporting is different from other sites in a statistically significant way. Statistical tools like R or SAS can be used to calculate a Z-score by comparing the observed adverse event rates at investigator sites against the expected rates, assuming a Poisson distribution, to identify statistically significant deviations in a multi-center clinical trial. Risk-based cutoffs for these Z-scores can be used to create a traffic light warning system for central monitors to decide when to trigger actions at sites based on their green, amber, or red status for such a metric. For example, if a site's AE reporting rate Z-score flips from a green level to an amber level, a training about adverse event reporting for site personnel in addition to a closer review of adverse events by medical professionals may be warranted. If a site's AE reporting rate Z-score flips to red, a more in-depth on-site audit of source records may be required. Identifying and remediating issues such as under- or over-reporting of AEs by sites as early as possible not only helps with collecting quality data, but also increases patient safety. Such a statistical approach can also extend to rates of patient discontinuation, clinically significant laboratory values, data discrepancy queries, data changes, and other metrics using investigator sites or countries as the units of analysis.

PATIENT PROFILES

When combined with trend analyses, Patient Profiles allow medical professionals to deep-dive into detected signals and trends by reviewing efficacy and safety data at the patient level. Visualizations help to depict patterns in data over tim and to reveal outliers. For example, a reviewer may miss a key insight when reviewing separate data listings of concomitant medications, adverse events, laboratory results, and vital signs. By visually combining the start and end dates of AEs, medications, laboratory results, and vital signs, a medical reviewer can quickly identify a potentially relevant medical interaction. By adding 95% confidence intervals from the clinical trial population to individual patient results, a reviewer can also make a determination when to investigate an individual result as a potential clinical outlier or data discrepancy.

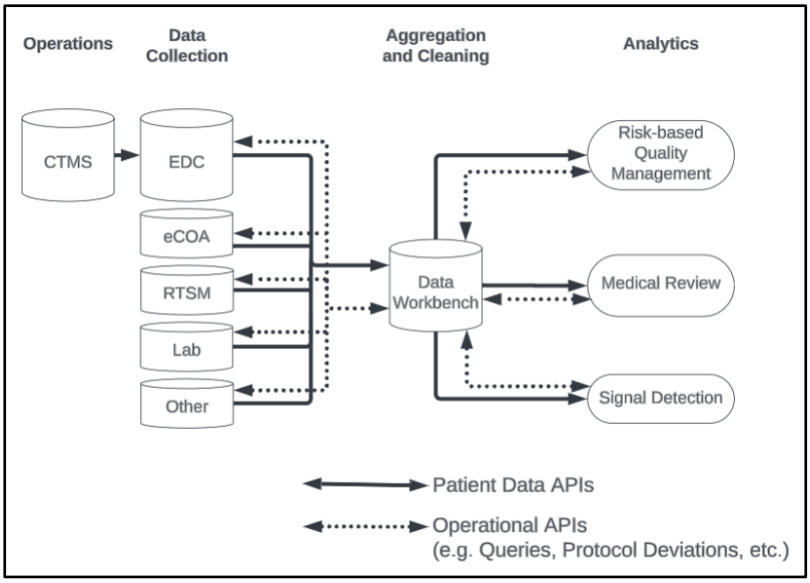
When used in concert on a clinical trial, data workbenches and analytics tools enable data managers, medical reviewers, and other functional experts to maximize the efficiency and effectiveness of their individual efforts. However, these individual benefits can be greatly magnified if the fit-for-purpose tools used by different functional groups are tightly integrated to minimize duplication of effort and maximize quality and speed of action.

INTEGRATING MODERN SYSTEMS WITH APIs FOR GxP COMPLIANCE

Modern APIs facilitate seamless communication between systems, reducing manual effort and improving efficiency. The availability for such connected applications turns traditional static, offline reports into data products that have processes and workflows built into their design. Some data products, such as clinical data

workbenches, are designed to be applicable across a broad array of organizations, since all of these organizations require data to be aggregated, reconciled, and cleaned. Other data products, such as those configured using open source data science toolkits, can be more bespoke in nature. These specialized data products can fit within specific organizational processes that may vary between large biopharma and small biotech, or organizations with integrated centralized monitoring and data management functions vs. independent departments.

FIGURE 1
Connected data products enable real-time data review and interact with productized workflows via APIs.



For example, a medical reviewer looking at a disconnected, static graph in a PDF file would need to document their observations in an email or spreadsheet tracker to be shared with a clinical data manager. This process is highly effort intensive and difficult to follow consistently. Instead, a medical reviewer using a constantly up-to-date visualization tool such as an R Shiny app can send a data discrepancy query back to the source EDC system via API, eliminating the need for manual query entry. As another example, a clinical data manager looking at an offline data listing produced by SAS might make a list of possible protocol deviations in a spreadsheet to be shared with a clinical operations team member. This effort is both error-prone and difficult to track to resolution. Instead, a data workbench can be used to automatically detect and flag potential protocol deviations like out-of-window assessments, creating candidate records in the CTMS via API. Finally, monitoring teams reviewing tabular operational reports and individual patient casebooks may struggle to consistently identify site-level performance issues or risks to patient safety. But the same team using a connected, constantly refreshed RBQM tool can spot statistically significant differences in data discrepancy trends, protocol deviations, or clinical assessments between investigator sites and trigger a Key Risk Indicator (KRI) back in the CTMS.

Users of data workbenches and analytics tools benefit from utilizing robust, GxP-compliant workflows. These workflows reduce manual tracking of tasks by automating documentation and audit trails. These audit trails help to guarantee that regulatory inspectors can reconstruct all actions performed within and across tools. A connected ecosystem of CTMS, EDC, data workbench, and analytics tools can provide a complete history of activity across a clinical trial, allowing end-to-end traceability of clinical and operational data.

CONCLUSION

Data workbenches and analytics tools each play essential roles in clinical trial data management. By leveraging the strengths of these tools and integrating them through modern APIs, organizations can optimize workflows, enhance data quality, and enable insightful analyses. A risk-based approach to data cleaning, supported by upstream data workbenches, ensures that analytics tools operate efficiently, delivering insights for decision-making. Together, these tools represent the future of streamlined, compliant, and effective clinical trial operations. The call to action for data scientists is to understand how to utilize the features and capabilities of each tool to enable clinical trial professionals to be maximally effective at their jobs.

REFERENCES

"Good Statistical Monitoring {gsm}: An R Package for Risk-Based Monitoring and Clinical Trial Analytics." Phuse US Connect 2024. https://www.lexjansen.com/phuse-us/2024/ar/PAP_AR04.pdf

“Clinical Data Workbenches: A Buyer’s Guide.” Veeva Systems. https://www.veeva.com/resources/clinical-data-workbenches-a-buyers-guide

“A 'Shiny' New Perspective: Unveiling Next-Generation Patient Profiles for Medical and Safety Monitoring.” https://www.lexjansen.com/pharmasug/2024/DV/PharmaSUG-2024-DV-382.pdf

CONTACT INFORMATION

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

Author Name: Casey Higgins
Company: Atorus Research
Email: Casey.Higgins@atorusresearch.com
Website: https://www.atorusresearch.com

Author Name: Pavel Burmenko
Company: Veeva Systems
Email: Pavel.Burmenko@veeva.com
Website: https://www.veeva.com