Paper AR07

Risk-Based Monitoring in Clinical Trials with Machine Learning Powered Anomaly Detection

Ece Kavalci, Lindus Health, London, UK
Zaid Al-Jubouri, Lindus Health, London, UK

ABSTRACT

Risk-based monitoring (RBM) is a critical component in clinical trials, ensuring data quality, patient safety, and regulatory compliance. In this conference, we present an innovative feature added to our clinical trial management platform that employs anomaly detection techniques for enhanced RBM in clinical trials. Our approach is to identify anomalies in trial data using machine learning, which may indicate potential risks that cannot be detected by hard-coded rules. This enables continuous monitoring and early risk mitigation, improving trial outcomes.The contribution of our research is the development of a robust, automated RBM system integrated within the clinical trial management platform, which efficiently identifies and addresses potential issues in clinical trials. We will discuss the challenges of traditional RBM and the methodology of our approach using a demo of this detection model integrated in our trial platform Citrus.The integration of an automated anomaly detection feature into the clinical trial management platform represents a significant advancement in the field, with the potential to enhance how clinicians monitor and mitigate risks in clinical trials.

INTRODUCTION

Clinical trials are highly regulated and complex processes that require careful coordination between all stakeholders. The burden of monitoring from a regulatory perspective falls on the sponsors as they are responsible for protection of participants and delivery of high-quality studies.

The regulations on effective monitoring are not specific about how sponsors are to conduct such monitoring so approaches have varied and depend on the type of trial being delivered. Historically research groups have relied on intensive monitoring methods like 100% source data verification and on-site visits to check data quality at source of collection. For various well documented reasons, this approach has become considered slow, inefficient and sometimes counter to the objectives of trials. From a non-regulatory perspective, this approach has been a contributor to slower trial delivery preventing faster delivery of treatments to patients.

To aid industry, FDA, MHRA and other regulatory agencies recommend using “Risk Based Monitoring” in trials using technology where the methods are appropriate under the circumstances of trials. The rationale behind this recommendation is that by focusing on risks around the most critical data and processes it is more likely to ensure subject protection and overall study quality compared to on site visits.

The key insight learnt from regulatory guidance to the industry is that a statistical approach to central monitoring can “help improve the effectiveness of on-site monitoring by prioritizing site visits and by guiding site visits with central statistical data checks”[1]. Work described in this abstract attempts to develop a novel approach to conducting this type of centralized monitoring in studies where clinical data is being collected in a single trial delivery platform.

TRADITIONAL RBM: CHALLENGES AND LIMITATIONS

Industry's adoption of risk based monitoring (RBM) approaches have been varied. On the simplest level, RBM can mean maintenance of a Risk Assessment and Categorization Tool (RACT) with a decision made to allocate risk proportional resources to monitor data and processes - herein called Light-RBM[2]. While this method adheres to the principles of RBM, it lacks the ability to, or does not fully leverage statistical tools to conduct analysis on top of these data and processes. Findings from a Light-RBM can help study teams to focus resources on higher risk items but not necessarily equip the team with tools to conduct complex oversight.

On the other hand, the current state of the art approach has been developed by CluePoints® who built a complex infrastructure that runs checks on incoming clinical data and provides visibility over Key Risk Indicators (KRIs) with pre-defined Quality Tolerance Limits (QTLs), herein referred to as the KRI-QTL approach[3]. This approach is included in the FDA and MHRA's guidance to the industry as a recommended approach to the industry. It's also aligned with TransCelerate® 's contributions to wider awareness of best practices in KRI monitoring.

This paper argues that approach is limited in its ability to deliver a truly dynamic and adaptive approach to risk based monotring. There are two key challenges identified:

- KRI's are tracked at study level and are retrospective by design. On any given trial, the delay between issues in a study emerging and being identified to aid intervention can have detrimental effects on study quality and participant safety
- While an effective tool for study manager and monitors across all participants and sites, this approach does not equip on-the-ground monitors and data-managers with real time feedback from a trial.

An example is provided here to demonstrate:

- **KRI:** Variance in Vital Signs Data per Site/Country/Study
- **QTL:** Cumulative Frequency in the study
- **Issue**: Participants in some sites have a high rate of vital sign reports that fluctuate heavily from visit to visit.
- **Cause**: Input validations for two different units share an inclusive zone - some users accidentally pick the wrong unit and the system doesn't block entry
- **Action**: Coordinate with stakeholders to change edit-check configuration to prevent further issues. Coordinate with data management teams to query and account for existing issues. Communicate to the biostatistics team to account for inconsistencies.

In this example we identify four limitations:

- **Feedback Speed:** The KRI/QTL configuration would in theory spot this issue once the frequency of occurrence crosses a threshold, which would necessarily be after a certain amount of time rather than instant feedback
- **Follow-Up Action:** The study team could now conduct an investigation into the causes and implement changes in the study processes or configuration to remediate the problem but the end user of the KRI/RBM tool does not have readily available access to best recommendations on follow-up
- **Access to information**: The current state of RBM tools lack an essential element - the review tool being directly integrated with the data collection methods. This will cause a gap between identifying the issue and digging through the records to see all occurrences of issues.
- **User experience:** Following from two limitations above, the lack of integrated tools makes it difficult to use these tools to navigate to the source of the problems and get full context

Given the above, the underlying argument of this paper is that there can and should be a way to prevent KRI's reaching unfavorable tolerance limits.

INTEGRATED APPROACH TO RISK BASED MONITORING

The approach described in the paper below is predicated on a novel approach to risk based monitoring.

We start by distinguishing three components to a system

- Clinical Database
- Participant specific Issue tracking
- Study/Country/Site scoped KRI tracking

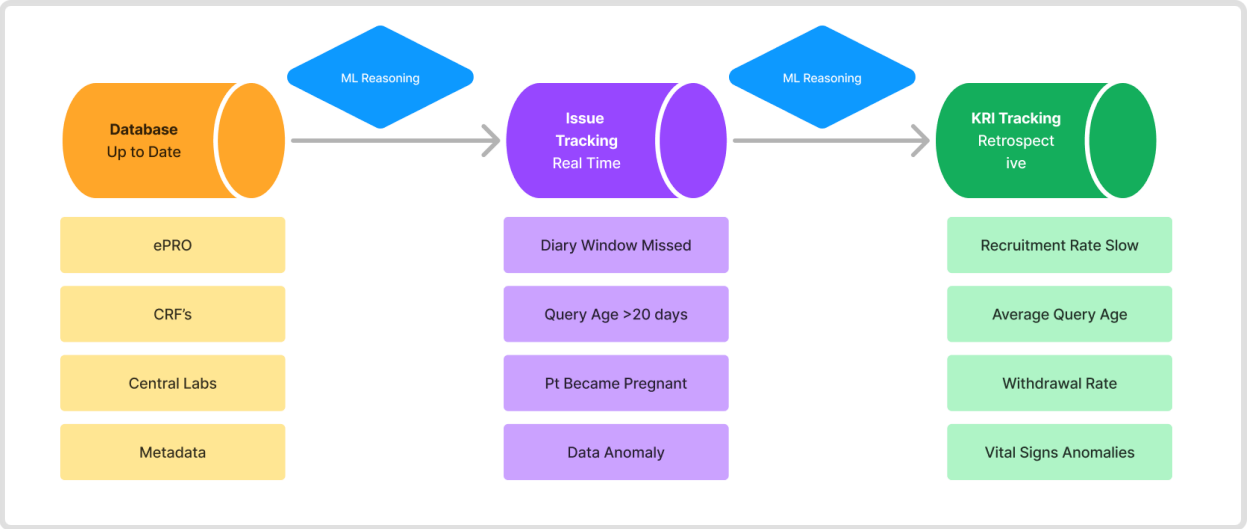

Figure 1 Clinical Database, Issue and KRI Tracking Complete workflow

The illustration above does not claim to introduce a new concept but a new approach. It identifies the following constraints:

- Database is an up-to-date and real time representation of data recorded in the systems
- Issue Tracking is a real-time (every 15 mins) consolidated issue tracking interface for coordinators and data managers to oversee conduct of the trial in real time and respond to issues in real time.
- KRI Tracking is a retrospective and a delayed (every 7-14 days) interface for study managers and sponsors for monitoring risks in a trial.

The ultimate goal of the approach is to ensure high-quality data and patient safety. The objectives are:

- Collect all data in a centralized, end-to-end tool and avoid the need for multiple systems to communicate
- Use machine learning (on top of other automated and manual processes) to maintain a real-time log of issues in the study with accountability for each
- Use machine learning and statistics to compute the KRI/QTL scores in real-time

This paper aims to demonstrate this approach in action, with a focus on how to use machine learning to detect anomalies in a specific type of data. The broader argument is that once these machine learning algorithms are incrementally implemented, the net-gain for quality of monitoring will be larger than the sum of parts of the various parts of the approach. The same approach described here can be used on all other types of issues in a clinical trial.

ANOMALY DETECTION IN MACHINE LEARNING

Anomaly detection, also referred to as outlier detection, is detection of unusual patterns in data which helps to identify any hidden error, problems or new behaviors. It is a crucial process for various domains including healthcare, finance and security. Traditional methods towards outlier detection employ rule-based methods or thresholds to spot anomalies; on the other hand, machine learning based methods adaptively learn from the data which makes them more robust and adaptive for unpredictable scenarios.

In Risk-Based Monitoring (RBM) within clinical trials, a primary advantage of machine learning-based outlier detection is the scalability of these algorithms. Traditional methods, which rely on set rules or thresholds, face difficulties when handling datasets with multiple variables. In our study, we focus on a dataset containing four variables from vital signs data. However, this approach can be extended to other types of data collected in clinical studies, some of which might include even more variables. Another key benefit is that machine learning models can recognize complex patterns in the data, a task that would be time-consuming and challenging using manual rule-based methods. Additionally, machine learning models can be easily adjusted and optimized using model settings or hyperparameters. This can automate much of the detection process, making it more efficient than traditional methods and speeding up the overall analysis.

METHODOLOGY: EMPLOYING ANOMALY DETECTION FOR RBM

Selected KRI is “Abnormal Trends in Data” in accordance with TransCelerate® Risk Indicator Library [2]. Within the scope of this project, only vital signs data is considered. Several anomaly detection approaches were tested including statistical methods. Isolation Forest model selected as the most effective machine learning algorithm to detect anomalies in vital signs data.

Category	Sub Category	Risk Indicator	General or Specific	Level of Scrutiny	Frequency	Threshold Basis	Scorecard	Weighting	Mitigation Actions
Data Quality	Data Trends	Repeated values (freq & rate)	General	All levels apply	Cumulative	Static, simple statistical algorithm	No	No	Yes
Data Quality	Data Trends	Variance in Vital Signs Data	General	Protocol, Country and Site	Cumulative	No specific threshold	Yes	No	Yes

Table 1 Selected Key Risk Indicator(KRI) from TransCelerate® KRI Library

This also is real-time changing as detected anomalies added to the issues tracking interface. The suggested approach is to use the net list of issues detected in the system to compute the KRI score. We account for anomalies in a sliding window approach, so we need to count anomalies as they appear for a total.

Simple possibilities:

- # of anomalies(issues) / time passed in study
- Avg anomaly issues per week or biweekly
- Any other KRI requested by the study team

Integration in a Single Platform

One of the important predicates to this approach is having all the data in one place removing the need for interoperability specifications. There is a path building a single platform that can achieve this. Below is an illustration of how the different pieces could connect.

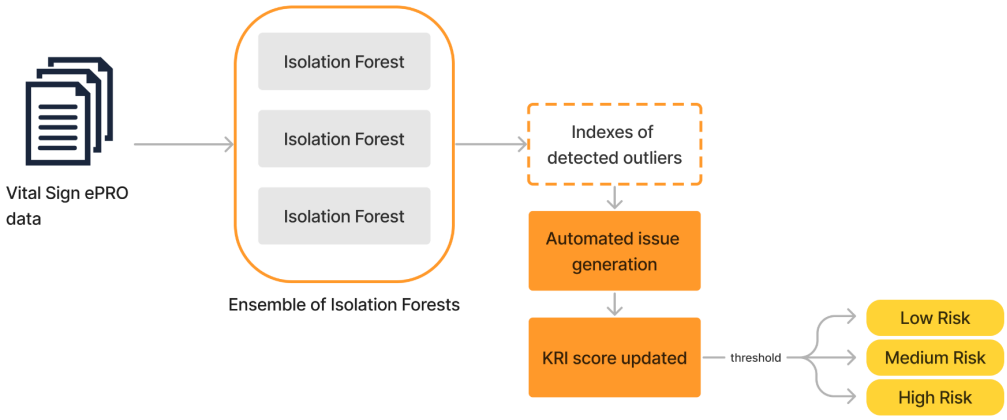

Figure 2 Anomaly detection KRI workflow chart

Isolation Forest Algorithm

Isolation Forest is an unsupervised learning algorithm designed for anomaly detection. It has a unique approach of isolating the data points compared to other outlier detection algorithms. Its core principle is that anomalies are data points that are few and distinct, making them more susceptible to isolation than the normal points. This algorithm operates by recursively partitioning the data space. It randomly selects a feature and then chooses a split value between the maximum and minimum values of that feature. This process continues until all data points are isolated or a predefined tree depth is reached. In the context of such isolation trees, anomalies are generally isolated nearer to the root, leading to shorter paths, while normal data points gravitate towards the deeper ends, entailing longer paths. The resulting anomaly score, based on these path lengths, provides a robust measure to spot outliers, with shorter paths correlating to a higher likelihood of a data point being anomalous.

Isolation forest does not use distance or density measured to detect the anomalies in the data which makes it a great fit for a dataset like this where the value range of the data is varying from participant to participant. Other algorithms would require normalization of the data to perform in a similar manner. However the variability in parameters like weight and blood pressure for different participants could cause subtle variations to be overshadowed, potentially masking genuine anomalies. Isolation Forest, despite not being inherently designed for time-series data, was selected for its robustness in handling anomalies in multi-dimensional spaces. Our dataset, comprising various vital signs, can be seen as a multi-dimensional space where each dimension corresponds to a vital sign, rather than a strict chronological series. Additionally, the biweekly nature of the ePRO data means that the time-dependent structure might not be as pronounced as in high-frequency time series. Future work might require more in-depth research into time-dependent anomaly detection models for the mentioned time-dependent datasets.

DATA PREPARATION

The vital signs ePRO consists of weight, waist circumference, systolic blood pressure and diastolic blood pressure questions. These surveys are sent to participants biweekly and in total every participant is expected to have 13 data points where each includes the 4 types of vital signs value and selected unit types. The ePRO surveys have hard validation to prevent unrealistic values.

Vital Sign	**Weight**	**Waist Circumference**	**Systolic Blood Pressure**	**Diastolic Blood Pressure**
Unit	kg / lb / stones	cm / inch	mmHg	mmHg
Hard Validation	Specified by study	Specified by study	Specified by study	Specified by study

Table 2 Vital Sign ePRO data type details

Preprocessing applied on study data which is in tabular format. Each row consists of the vital sign type, result in original unit, original unit, standard unit, result in standard unit and the anonymised subject id. For all the numerical data to be in the same scale, only the unified unit and calculated vital sign values were used for the anomaly detection. In order to be able to apply multivariate algorithms, the dataset pivoted to have values from each survey for each participant as a row.

Vital Signs Test Name	Result or Finding in Original Units	Original Units	Numeric Result/Finding in Standard Units	Standard Units	Study Day of Vital Signs	SubjectID
Weight	129.85	kg	129.85	kg	19	sid-010
Waist Circumference	51.00	in	129.54	cm	19	sid-010
Systolic Blood Pressure	156.00	mmHg	156.00	mmHg	19	sid-010
Diastolic Blood Pressure	86.00	mmHg	86.00	mmHg	19	sid-010
Weight	123.00	kg	123.00	kg	33	sid-010

Figure 3 Simplified vital signs dataframe

The anonymised vital signs data represent study data and participants had started the study in different time points. Hence, the number of submitted ePRO surveys and study day of the submission will differ for each participant. To accommodate these differences, an instance index added to the dataset instead of using actual recorded study day fields.

SubjectID	Diastolic Blood Pressure	Systolic Blood Pressure	Waist Circumference	Weight	instance_index
sid-083	78.0	122.0	134.0	146.0	1
sid-083	65.0	105.0	127.0	144.0	2
sid-083	75.0	115.0	135.0	144.0	3
sid-083	68.0	118.0	131.0	143.0	4
sid-083	67.0	115.0	136.0	141.0	5

Figure 4 Pivoted vital signs dataframe

The latest version of the dataset presented in this paper contains 523 rows in the pivoted dataset. One row is eliminated from the final dataset due to having a missing value for Waist Circumference. This information feeds into another KRI but this is not in scope of this paper.

MODEL IMPLEMENTATION AND EVALUATION

Computational Efficiency

A technical challenge to be overcomed in this project is the computational efficiency of real-time machine learning based anomaly detection. A real-time approach towards this KRI means the algorithm will run everytime a new reading is added to the dataset in hand. In order to keep the outlier detection computationally effective two approaches were selected to be tested. One of these is to use sliding windows where the window size is approximately half of the maximum instance size for each participant. For this dataset window size is selected as 6 as the total expected number of readings is 13. In this method last n(window size) readings for each participant selected to feed into the outlier detection algorithm.

In the context of this project, where each participant has a limited number of readings, up to a maximum of 13, an alternative to the sliding window approach is proposed. By adjusting the maximum sample size parameter in the Isolation Forest to half the dataset size, each tree in the algorithm is trained on a randomly sampled subset that encompasses approximately half of all available data. This randomness presents two main benefits over a rigid, sequential windowing approach. First, the random selection provides each tree a varied perspective, ensuring diverse and comprehensive coverage of potential anomalies. It reduces the risk of missing outliers that might be at the boundary of a strict window. Second, it offers computational efficiency by eliminating the need for manual data segmentation and repeated model training on various segments. Overall, this method optimally balances computational speed with robust anomaly detection, particularly apt for datasets with a limited range of readings like the one under consideration.

EVALUATION OF RESULTS AND DISCUSSION

In this project, we used a straightforward yet effective method for evaluation: we visually plotted the data points. By looking at the graphs of the vital signs data, it's easier to spot unusual patterns or outliers. We then compared what we saw in these plots to the results from the Isolation Forest algorithm. This approach helped us see how well the algorithm's findings matched up with clear visual trends in the data. Using graphs for validation is especially useful in health-related datasets because it ensures that the anomalies the algorithm detects make sense in a real-world, medical context. Simply put, this method makes sure that the anomalies we identify are both mathematically and medically important.

As can be seen in the plots, most of the anomalies detected are due to the waist circumference data. The detected anomalies which show a normal distribution in the waist circumference plot shows anomalies in the other plots. Out of a total of 522 readings, the applied algorithm identified 15 anomalies. When cross-referenced with the ground truth data, 9 of these detections were confirmed as genuine anomalies, resulting in a precision of 60%. The dataset contains 11 true anomalies, and the algorithm captured 9 of them, leading to a recall or sensitivity of approximately 81.82%. The F1 score, a balance between precision and recall, is calculated to be roughly 69.23%. In the context of study-critical data, a high recall is especially paramount. While precision is important, ensuring that a majority of critical anomalies are detected (even at the expense of some false positives) is essential to maintain the integrity and safety of the study. The results validate the algorithm's capacity to capture most of the genuine critical anomalies, though further optimization might reduce the number of non-critical readings flagged.

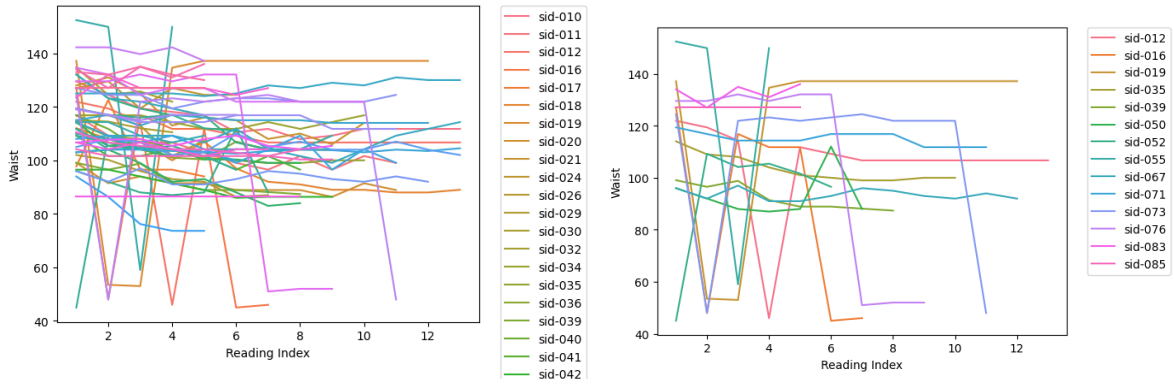

Figure 5 Waist circumference data plot of all participants vs anomaly instances data plot respectively

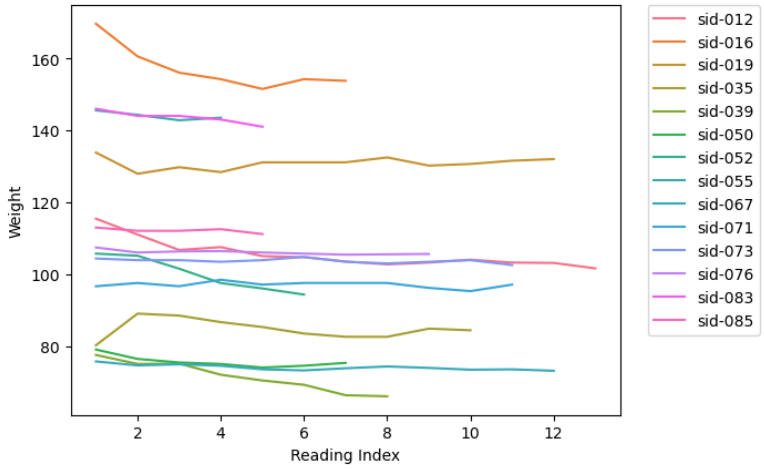

Figure 6 Weight data plot for anomaly instances

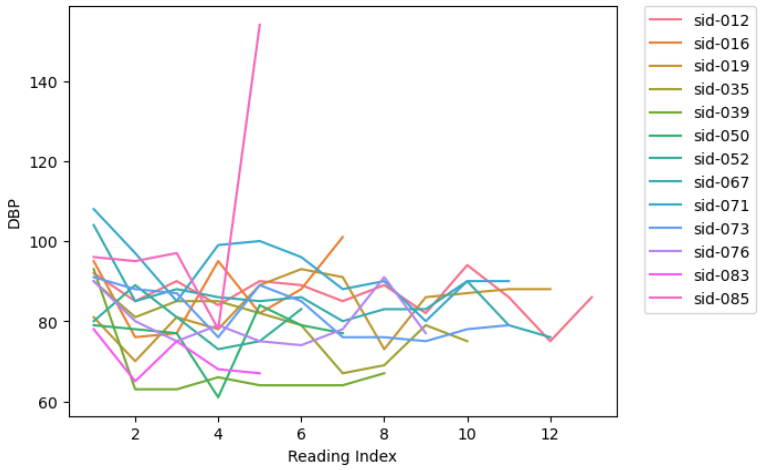

Figure 7 Diastolic Blood Pressure data plot for anomaly instances

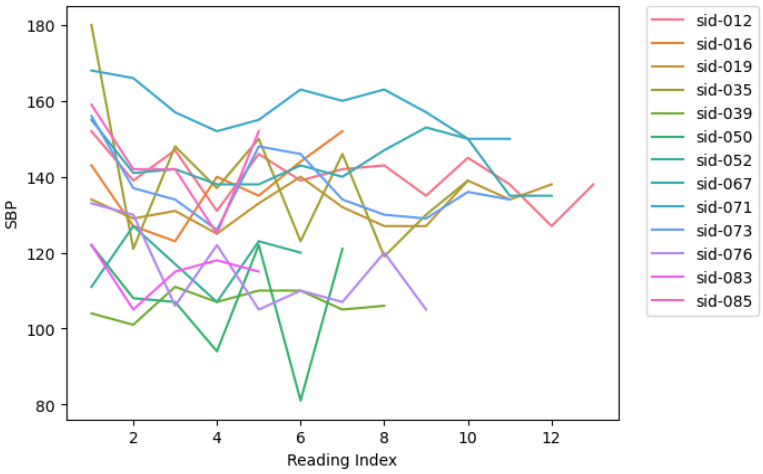

Figure 8 Systolic Blood Pressure data plot for anomaly instances

BENEFITS OF THE INTEGRATED RBM SYSTEM

The technical implementation described here describes a method for systematically detecting anomalies in data and flagging these to data managers/monitors in real time. While this approach can be successfully implemented using multiple systems, we think it's crucial that the different pieces of this are all within a single location.

There are some serious challenges to implementing this in practice and using in real clinical trials - the system has to be reliable and foolproof and have sufficient functionality to be feature complete for effective monitoring. Traditionally these challenges would motivate using different specialized tools to make this approach work. We think that while there are benefits of running more complex trials by using multiple specialized software, the net benefits gained by a successful implementation of these methods in a single platform, has potential for much larger impact over time than trying to patch different tools together.

CONCLUSION

In this paper, we introduced a machine learning-based anomaly detection feature for enhanced Risk-Based Monitoring (RBM) in clinical trials, emphasizing its ability for continuous monitoring and early risk detection. Traditional RBM, while effective at the study level, has limitations in adaptability and timeliness, which this new system addresses. The continuous approach offers more prompt identification of potential risks, leading to better trial outcomes, cost-efficiency, and patient safety. Furthermore, integrating this feature into a single platform simplifies the monitoring process by eliminating the need to juggle multiple systems, ensuring more streamlined data management and user experience in clinical trials.

REFERENCES

1. U.S. Food & Drug Administration. (n.d.). A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/risk-based-approach-monitoring-clinical-investigations-questions-and-answers
2. TransCelerate BioPharma Inc. (n.d.). Risk Based Monitoring Solutions. Retrieved from https://www.transceleratebiopharmainc.com/assets/risk-based-monitoring-solutions/
3. Wolfs, M., Bojarski, Ł., Young, S., Cesario, L., Makowski, M., & Sullivan, L. B. (2023). Quality Tolerance Limits' Place in the Quality Management System and Link to the Statistical Trial Design: Case Studies and Recommendations from Early Adopters. *Therapeutic Innovation & Regulatory Science, 57*(839–848). [Open access]. https://link.springer.com/article/10.1007/s43441-023-00504-6

CONTACT INFORMATION

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

Ece Kavalci
Lindus Health
London
ece@lindushealth.com
https://www.lindushealth.com/