Paper IC03

Synergizing Large Language Models (LLMs) and Human-in-the-Loop (HITL) for Data Noise Reduction in the RBQM approach

Mukesh Babu K, Nimble Clinical Research, Bangalore, India
Vineet Jain, Nimble Clinical Research, New Jersey, USA

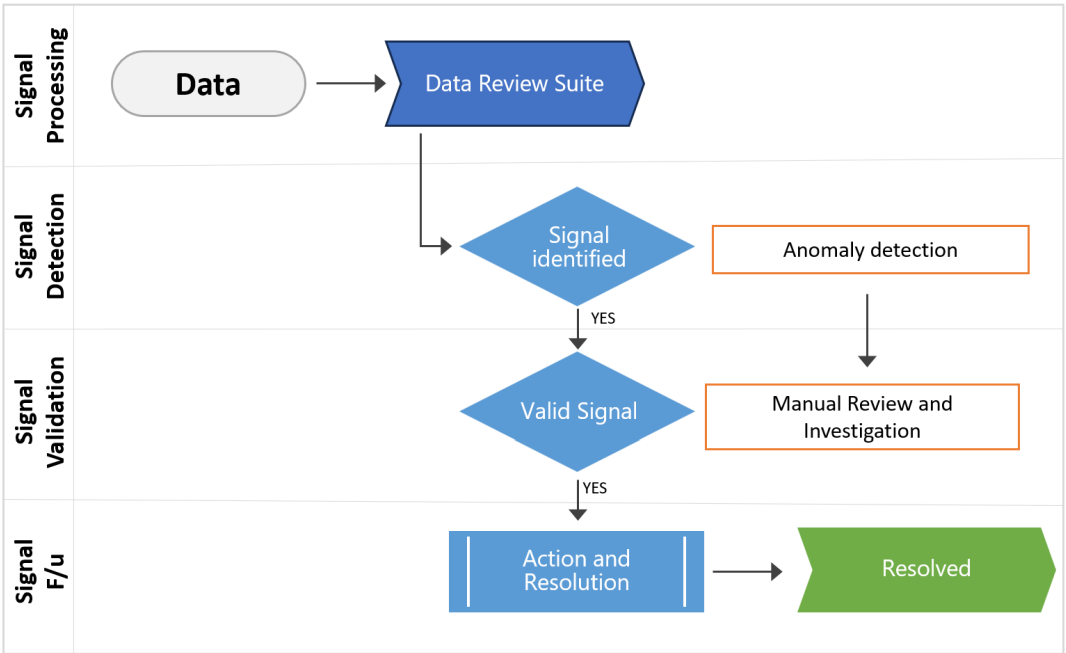
ABSTRACT

In Risk-Based Quality Management (RBQM), the pursuit of signal validation amid data noise is critical for its success. Noisy signals can make detecting valid signals of safety or quality issues challenging. Dealing with noise can be time-consuming and resource intensive. It may require additional data cleaning and analysis efforts, slowing the RBQM process and increasing operational costs. On the other hand, applying LLM techniques in RBQM is an unexplored territory. As a tool, LLM offers promising avenues for distinguishing meaningful signals from all data signals. By leveraging LLM with the Human-in-the-loop (HITL) approach, RBQM practices can enhance their ability to identify relevant patterns and insights, ultimately improving the quality and efficiency of clinical trials. Given the persistent challenge of data noise, a refined approach to signal validation is imperative in RBQM. This abstract highlights ongoing efforts to leverage LLM & HITL for effective RBQM and signal validation. It highlights its potential to reduce noise and enhance decision-making precision in RBQM success.

SIGNAL VALIDATION IN RBQM

Risk-Based Quality Management (RBQM) is an approach to managing the quality of clinical trials in a more targeted and efficient manner. RBQM involves the identification, assessment, and management of risks throughout the entire clinical trial process. Signals are potential sources of risks that prompt further investigation and are followed by corrective actions.

Figure 1 details the current Signal detection and validation process in the RBQM approach.



Steps involved in traditional Signal Validation to reduce data noise:

Step 1 - Data Processing: Aggregated data from various sources are reviewed, utilizing analytical tools and statistical methodologies to identify signals.

Step 2 - Detecting Signals: Clinical and operational data is systematically reviewed at predefined intervals, using statistical methods and analytics tools to identify trends, patterns, outliers, and data inconsistencies to surface signals. These signals indicate potential risks.

Step 3- Signal Investigation: Signals are investigated further for building context (Supporting data, possible root cause, suggested action/mitigations) using the knowledge and experience of data reviewers. This involves a manual review to ensure that the signal is real and not a false positive.

Step 4 - Follow-up on Signals: Actionable signals and associated risks are communicated to the respective action owners for corrective and preventive measures.

This process will be followed and repeated at every cycle of review.

CHALLENGES IN THE CURRENT APPROACH TO DATA NOISE

In the RBQM approach, risk signals may have data noise originating from various sources, such as measurement errors and signal interference, impacting signal validity and quality. While traditional techniques like filtering and predefined rules can be employed during data preprocessing to address data noise, their effectiveness is not guaranteed, as they may impede signal identification. An alternative approach involves signal validation through context building by data reviewers, i.e., looking at related domains to substantiate the signals, which proves effective but is labor-intensive and manual, posing challenges to signal validation. Therefore, there is a pressing need for advanced methodologies in the signal validation process.

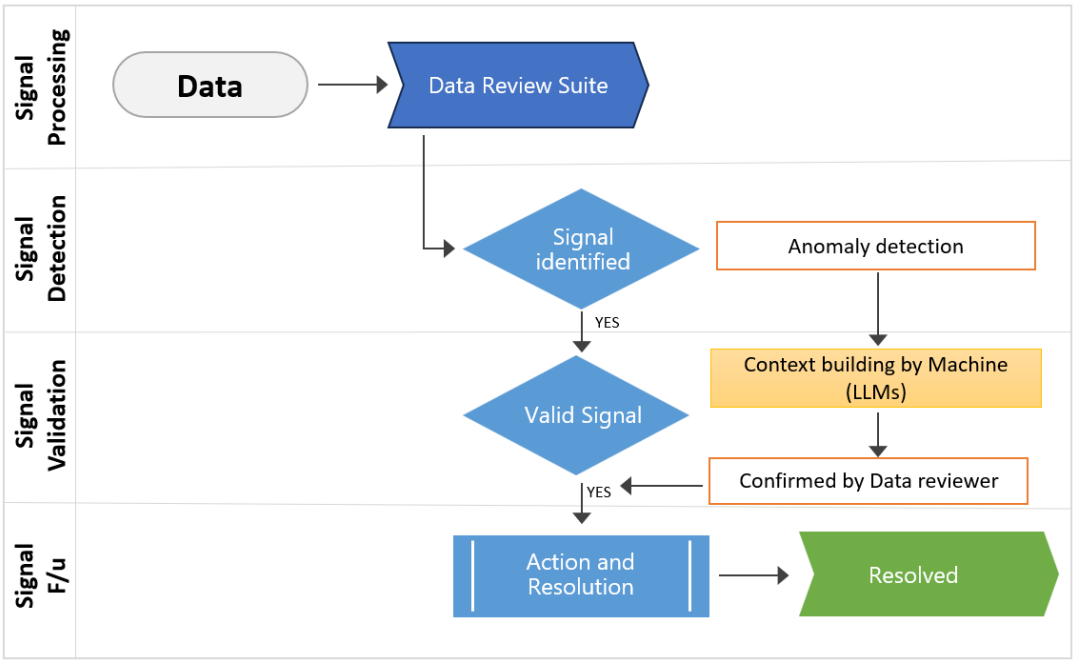
Below are the limitations of current Noise reduction practices:

- Traditional filtering limitations: Filters might remove noise and valid signals, especially when dealing with complex or nuanced data.
- Predefined rules' inflexibility: Rigid rules may fail to adapt to changing patterns or new risk scenarios, leading to missed signals.
- Context building's labor intensity: Researching related domains for every signal is time-consuming and impractical for large datasets.
- Subjectivity and bias: Human reviewers' interpretations and individual biases can influence the validation process and its outcomes.

USE OF LLM SIGNAL VALIDATION TO REDUCE NOISE

Utilizing large language models (LLMs) revolutionized clinical trials and introduced a new era of enhanced signal detection and streamlined processes. We have uncovered unexplored territory in the field by leveraging LLMs, specifically exploring context building through data summarization for signal validation. This innovative approach aims to identify substantial evidence around signals, providing a promising avenue for improvement.

Figure 2 details the current Signal detection and validation process in the RBQM approach.



Steps involved in LLM- HILP Signal Validation to reduce data noise:

Step 1 - Data Processing: Aggregated data from various sources are reviewed, utilizing analytical tools and statistical methodologies to identify signals.

Step 2 - Detecting Signals: Clinical and operational data is systematically reviewed at predefined intervals, using statistical methods and analytics tools to identify trends, patterns, outliers, and data inconsistencies to surface signals. These signals indicate potential risks.

Step 3a - Investigating Signals: Utilizing LLMs, i.e., context-based input prompts and summarization/Corroboration/Tracking are generated from all pertinent data domains.

Step 3b - Validation by Data Reviewer: The output from LLMs is then scrutinized and validated by the Data Reviewer, prompting further follow-up and action.

Step 4 - Follow-up on Signals: Actionable signals and associated risks are communicated to the respective action owners for corrective and preventive measures.

CONTEXT BUILDING BY LLMs – POSSIBLE SCENARIOS:

We have selected site-level KRI for AE Rate (number of AEs/per subject visit) and presented three different use cases indicating how LLM can benefit the data reviewers in signal validation.

1. Context Building by Data Summarization

The most direct application of LLMs in signal validation is the summarization of related data used to accept/reject the signal as noise. This approach suggested paths for investigating the KRI by selective summarization. The selective summarization capabilities of LLMs provide quick insights into multi-dimensional trial data. This approach enhances the understanding of AE data and supports more effective and timely decision-making in clinical trial management.

To investigate AE rates, we can ask LLMs to look at the below data of interest to provide summarization:

- Subject Demographics and AE Patterns: The summary could illustrate the subject age, gender, or pre-existing conditions, frequent AEs, and the severity of AEs, enabling potential identification of patterns.
- Other KRIs such as Concomitant medication Rates & Medical History rates: LLM can summarize other key data points, providing a clear picture of other related data points and corresponding trends showing levels of other KRIs. For instance, it might highlight that higher concomitant medication rates can be linked to a higher frequency of certain AEs.

LLM prompt:

Create a summary of the following site-level demographics and Adverse event aggregations.

Subject Demographics: Age Groups: 18-30: 25% 31-45: 35% 46-60: 20% 60+: 20% Gender Ratio: Male: 40% Female: 60%.

Pre-existing Conditions: Diabetes: 15% Hypertension: 20% cardiovascular diseases: 10% ……

Frequency of AEs: Nausea: 30% Headache: 25% Dizziness: 20% Fatigue: 15% Rash: 10% ….

Severity of AEs: Mild: 40% Moderate: 35% Severe: 25%

AE Rate at This Site: 5 Average (SD) AE Rate Across Other Sites: 3 (0.5)

CM Rate at This Site: 11 Average (SD) AE Rate Across Other Sites: 8 (1.4)

MH Rate ….

Create one summary paragraph each for Subject Demographics, Adverse Event Patterns, & Site-specific Rate Comparison.

LLM's response to the above prompt:

Subject Demographics: *The demographic profile at this site is predominantly young and female, with the majority being under 45 years of age. Notably, a significant portion of subjects have pre-existing conditions like hypertension and diabetes, which are common in about one-fifth and one-sixth of the subjects, respectively.*

Adverse Event Patterns: *Adverse events reported predominantly include nausea and headache, with other significant occurrences like dizziness and fatigue. The severity of these events varies, but there is a considerable incidence of moderate to severe cases.*

Site-Specific Rate Comparison: *The site exhibits a higher AE rate than the average across other sites. Similarly, the Concomitant Modification Rate is elevated here, indicating a higher frequency and potential severity of clinical modifications in response to AEs at this site than others in the study.*

The prompt can be modified to provide a more detailed, concise, or insightful response. This example highlights how large amounts of data can be summarized in the desired manner with carefully designed prompts.

2. Context Building by Corroboration

In this use case, we are going further to find corroborative evidence to either validate the signal or reject it as noise.

LLM prompt:

Determine if the high Adverse Event (AE) rate at a specific site is valid or noise by analyzing the following key co-related metrics and comparing them to averages across other sites:

AE Rate: Site: 4.5 Other Sites Average: 3.0 (SD: 0.7)

MH Rate: Site: 3.8 Other Sites Average: 2.5 (SD: 0.6)

CM Rate: Site: 2.6 Other Sites Average: 1.9 (SD: 0.5)

AE Duplication Rate: Site: 0.5 Other Sites Average: 0.3 (SD: 0.1)

Data Entry Delay (days): Site: 2 Other Sites Average: 1.5 (SD: 0.4)

Average Exposure (days): Site: 80 Other Sites Average: 115 (SD: 8)

Usually, a high AE rate, accompanied by a high CM rate, MH rate, AE duplication rate, and. exposure, and low data entry delay, justifies a high AE rate. Assess if the site's high AE rate is an outlier considering these co-related metrics. Present the conclusion in 50-100 words, followed by a detailed analysis.

LLM's response to the above prompt:

*Conclusion: **The high AE rate at the site appears justified and not an outlier when considering the correlated metrics**. Elevated MH, CM, and AE duplication rates and a slightly increased data entry delay align with the higher AE rate. However, the notably lower average exposure at this site diverges from the typical pattern, suggesting that the high AE rate might be influenced by other factors.*

Detailed Analysis:

AE Rate: The site's AE rate of 4.5 is significantly above the average of 3.0 at other sites. The deviation is notable, considering the standard deviation 0.7 at other sites.

MH Rate: The MH rate at the site is 3.8, much higher than the 2.5 average at other sites. Given its standard deviation of 0.6, this elevated rate suggests a higher prevalence of medical history issues, which could contribute to the increased AE rate.

CM Rate: With a CM rate of 2.6 at the site versus an average of 1.9 at other sites, and considering the standard deviation of 0.5, there is a clear increase in clinical modifications, likely in response to more frequent AEs.

AE Duplication Rate: The site's AE duplication rate of 0.5, higher than the 0.3 average at other sites (standard deviation of 0.1), indicates more instances of repeated AE reporting, contributing to the higher overall AE rate.

Data Entry Delay: The site's average data entry delay of 2 days, slightly above the 1.5-day average at other sites (standard deviation of 0.4), points to marginally slower reporting, which could affect AE rate calculations.

Average Exposure (Days): The average exposure of 80 days at the site is significantly lower than the 115-day average at other sites (standard deviation of 8). This lower exposure duration contradicts the expectation that a higher AE rate correlates with longer exposure.

This example indicates the potential of Large Language Models (LLMs) with detailed prompts to gain complex insights swiftly and efficiently. We can rapidly assess and interpret intricate data relationships by feeding detailed, site-specific metrics into such prompts. Analyzing varied metrics like AE and MH rates provides a nuanced understanding of why a site might show a higher AE rate. Although LLMs can identify the correct correlation across multiple metrics, in our experiments, we have found it is better to identify positive/negative correlations upfront to obtain predictable responses.

3. Context building by Tracking

Finally, another prompt is provided below that further investigates the metric trend and previous communication on signal action/resolution to determine the continued management of a previously identified signal.

LLM prompt:

Review the progress of the following signal identified four weeks ago as part of RBQM central monitoring process based on the new data compared with previous data values and provide a conclusion in 50-100 words.

Previous Signal:

Given the recent analysis indicating an anomaly in your site's Adverse Event (AE) rate, we request an immediate investigation into potential causes. The AE rate at your site is 4.5, notably higher than the average of 3.0 across other sites. However, your site shows lower rates in Medical History (MH), Concomitant Modification (CM), and a similar AE Duplication rate compared to other sites, which contradicts the high AE rate. Please review your data collection and reporting processes.

Following are current KRI values for the site, previous KRI values for the site & current average. KRI value across sites after a month:

AE Rate: 3.5, 4.5, 3

MH Rate: 1.8, 1.8, 2.5

CM Rate: 1, 0.6, 1.2

AE Duplication Rate: 0.3, 0.5, 0.3

Data Entry Delay: 2,2, 3

Average Exposure: 110,85, 135

Note: Usually, a high AE rate, when accompanied by a high CM rate, MH rate, AE duplication rate, and. exposure, and low data entry delay, justifies a high AE rate.

LLM's response to the above prompt:

The reduction in the AE rate and improvements in the CM rate and the AE duplication rate suggest that the actions taken in response to the initial signal have been effective. The site's performance in these KRIs is more in line with the average across other sites, indicating a successful resolution of the initial anomaly. However, continued monitoring of the MH rate and average exposure is recommended to ensure alignment with broader trends and identify any emerging concerns.

This conversation exemplifies how careful prompt design can generate customized responses for every signal/noise classification stage in Risk-Based Quality Management (RBQM). An initial prompt was created to summarize the data. The subsequent prompt was tailored to initiate a site-specific investigation, request corrective actions, and eventually prompt to re-evaluate the site's performance after implementing changes. These prompts were designed to include updated data, enabling a comparison of new values against previous ones and averages across other sites. This approach highlighted improvements and areas needing continued monitoring, demonstrating the progress in resolving the initial anomaly. The careful crafting of these prompts at each stage of the RBQM process underscores the potential of using advanced analytical methods to derive tailored insights.

USE OF THE RIGHT TECHNOLOGY FRAMEWORK FOR LLM

A robust, secure, and flexible platform is essential to harness the capabilities of Large Language Models (LLMs). Such a platform must excel in aggregating and flexibly processing complex datasets to generate effective prompts automatically that enable LLMs to derive meaningful insights effectively. Integrating advanced data tools is crucial for feeding complex structured information to LLMs, facilitating a more context-rich analysis.

CONCLUSION

Identifying anomalies goes beyond simply detecting data points that deviate from the norm. Understanding the context in which events occur is crucial for accurately identifying true anomalies and avoiding false positives. Large language models (LLMs) offer promise in signal validation by generating insights from complex data. Further, LLMs need a sophisticated platform, and advanced data tools provide intricately structured information, enabling deeper contextual analysis. The synergy between a sophisticated LLM data platform and Human-in-the-Loop (HITL) is key to efficient anomaly detection in complex data environments. This synergy stands as the key to success in such scenarios. Besides RBQM signal validation, a driven context-building approach can be applied to other areas, including safety signal identification, statistical monitoring, and data cleaning.

CONTACT INFORMATION

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

Mukesh Babu Krishnamurthy
Head of Clinical Analytics
Nimble Clinical Research
991 US Highway 22, Suite 200
Bridgewater, New Jersey 08807 USA
Email: mukesh.babu@nimble-cr.com

Vineet Jain
Founder & CEO
Nimble Clinical Research
991 US Highway 22, Suite 200
Bridgewater, New Jersey 08807 USA
Email: vineet.jain@nimble-cr.com