Paper AR04

Applying Central Statistical Monitoring to find anomalies and address quality issues in clinical trial data

Wiktor Majewski, GSK, Warsaw, Poland

ABSTRACT

Finding anomalies and correcting errors in clinical trial data as the trial is ongoing is a way to ensure consistent high-quality data output, that is critical to an accurate safety and efficacy assessment. This paper describes how Central Statistical Monitoring (CSM) is employed in Risk-Based Quality Management (RBQM) operation framework. Our application of CSM aims to identify trial sites and individual subjects with potential data quality issues and takes into account both time series and distribution of clinical parameters and events - on site and subject level. The findings – which include atypical data patterns, outlying values and low variability of variables (e.g. vital signs and laboratory test results) – complement the central monitoring targeted reviews and review of operational metrics like Key Risk Indicators and Quality Tolerance Limits. Observations and recommendations of statistical monitors are presented to data management and core study teams during periodic CSM reviews.

INTRODUCTION

Ensuring the high quality of data is critical for clinical trial success and requires a quality management process in operation throughout the entire course of trial. In Risk-Based Quality Management (RBQM) model – which supports study teams in allocating resources in areas which bring the most significant value, Central Monitoring (CM) focuses on critical data and processes and supervises operational risks by monitoring Key Risk Indicators (KRI's) and Quality Tolerance Limits (QTL's), while clinical data quality is verified via data management queries and source data review at the sites. The role of Central Statistical Monitoring (CSM) is to perform the holistic statistical review of the clinical trial data with no distinction between critical and non-critical parameters. It is applied to identify unusual patterns, outliers, trends or atypical data distributions at sites, and detect anomalies that may impact trial data quality. CSM aims to identify potential risks and has been proven to be a cost-effective complement to other data management and monitoring techniques.

APPLYING CSM

Our approach to application of CSM in the RBQM framework involves periodic statistical review of clinical data, that complements regular reviews of operational metrics and targeted CM reviews. In scope of the analysis are potentially all variables collected from case report forms (CRFs), patient-reported outcomes (PROs) and laboratory test results. We use unsupervised method based on assumption that data from all centers should be comparable and statistically consistent. Statistical review cycle is initiated once the study meets set criteria for minimum amount of data required to perform robust statistical tests: have at least 5 sites with 10 or more subjects enrolled/randomized, or at least 10 sites with 5 or more subjects enrolled/randomized.

IDENTIFYING ANOMALIES

Anomaly detection is based on a series of statistical tests comparing distributions of variables from each site to all other sites. All tests that are relevant to the type of variable (continuous, factor, date) are performed for each site with at least 10 records of a variable. The highest value of negative logarithms of p-values from all tests for each variable is considered as variable score. The highest variable score for a site multiplied by modulation factor (reduction factor for data noisiness, i.e. adjustment for site size) constitute site anomaly score. Anomaly scores order the sites according to the likelihood of deviation of site's data from data at all other sites not being result of a chance, guiding the review towards the sites with most anomalous data.

$$\left.\begin{array}{l} v_1 \left/ \begin{array}{l} \xrightarrow{Test_{11}} p_{11}\ (p\text{-value}) \\ \vdots \\ \xrightarrow{Test_{1k}} p_{1k}\ (p\text{-value}) \end{array} \right\} score_{v_1} = \max\limits_{1 \le x \le k} (-\log p_{1x}) \\ \vdots \qquad\qquad \vdots \\ v_n \left/ \begin{array}{l} \xrightarrow{Test_{n1}} p_{n1}\ (p\text{-value}) \\ \vdots \\ \xrightarrow{Test_{nm}} p_{nm}\ (p\text{-value}) \end{array} \right\} score_{v_n} = \max\limits_{1 \le x \le m} (-log\, p_{1x}) \end{array}\right\} anomaly\ score = \max_{1 \le x \le n} \left(score_{v_x}\right) \times modulation\ factor$$

EXAMPLES OF OBSERVATIONS

Figures 1-3 show examples of anomalies identified by using CSM method described in this paper. Based on the site and site-variable scores alone it is not possible to determine the exact reason for the results. The data were highlighted by central statistical monitor and needed further reviews and checks to address the observations.

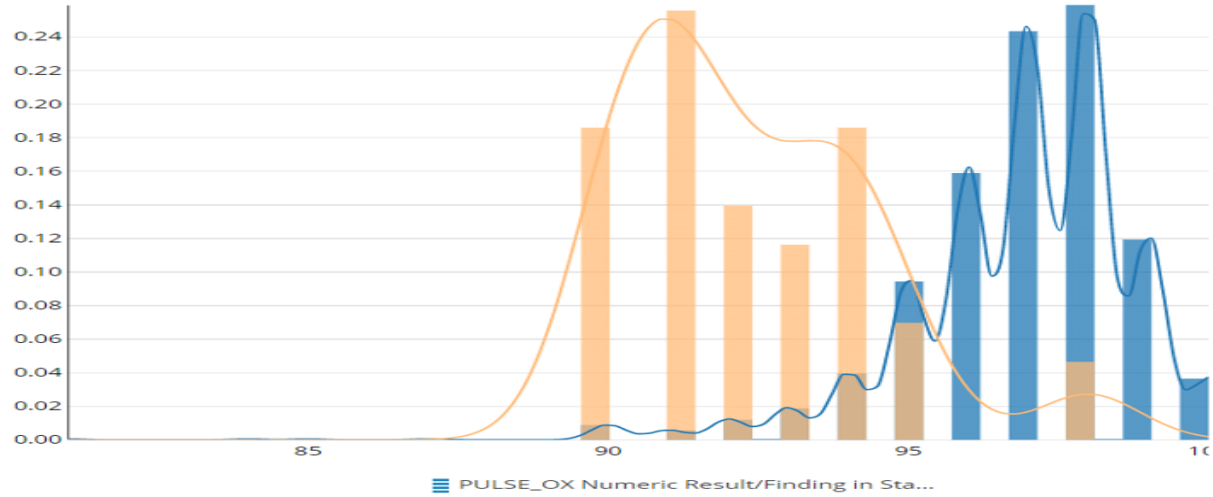

Figure 1. Example of CSM observation: Site A (amber) has atypical distribution of oxygen saturation [%] values: reported values that are in 95% within 90-95 range (median = 92) and are significantly lower than values from other sites in the trial (blue, median = 97).

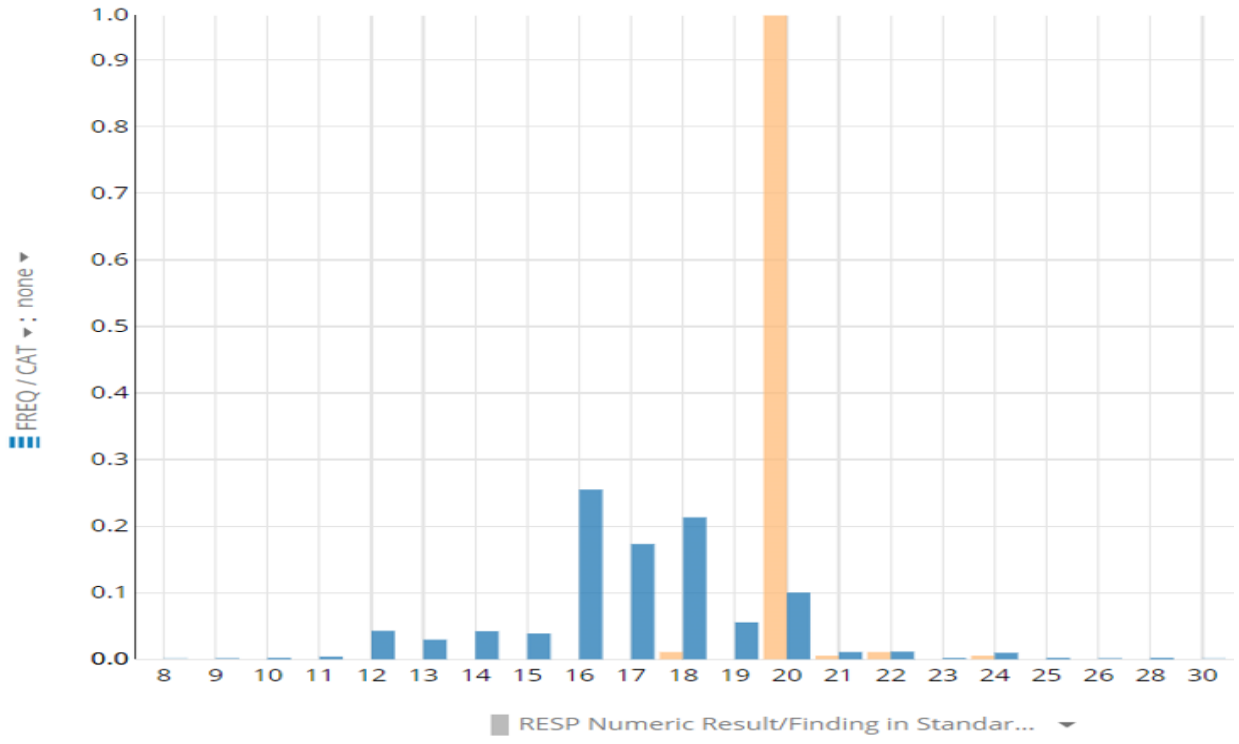

Figure 2. Example of CSM observation: Site B (amber) has low variability of respiratory rate measurement. Over 95% of all records at the site has a value of 20. The values from other sites in the trial (blue) are much more evenly spread across the range of values.

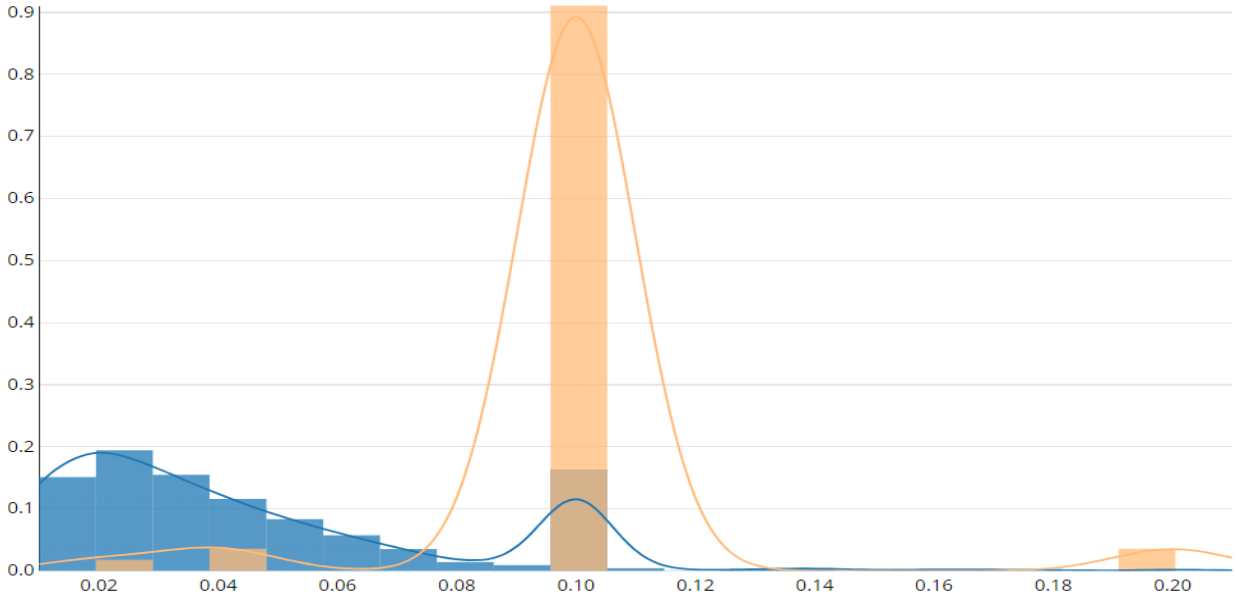

Figure 3. Example of CSM observation: Site C (amber) has recorded four specific values only. The values from other sites in the trial (blue) is distributed across the range with less preference towards specific digits/values.

Figures 4-6 present examples of cases where outlying values were identified. These type of observations are found at the site (or group of sites) as well as at individual subject level. Usually connected with data entry errors and unit conversion issues, they are likely to be observed within a group of associated variables.

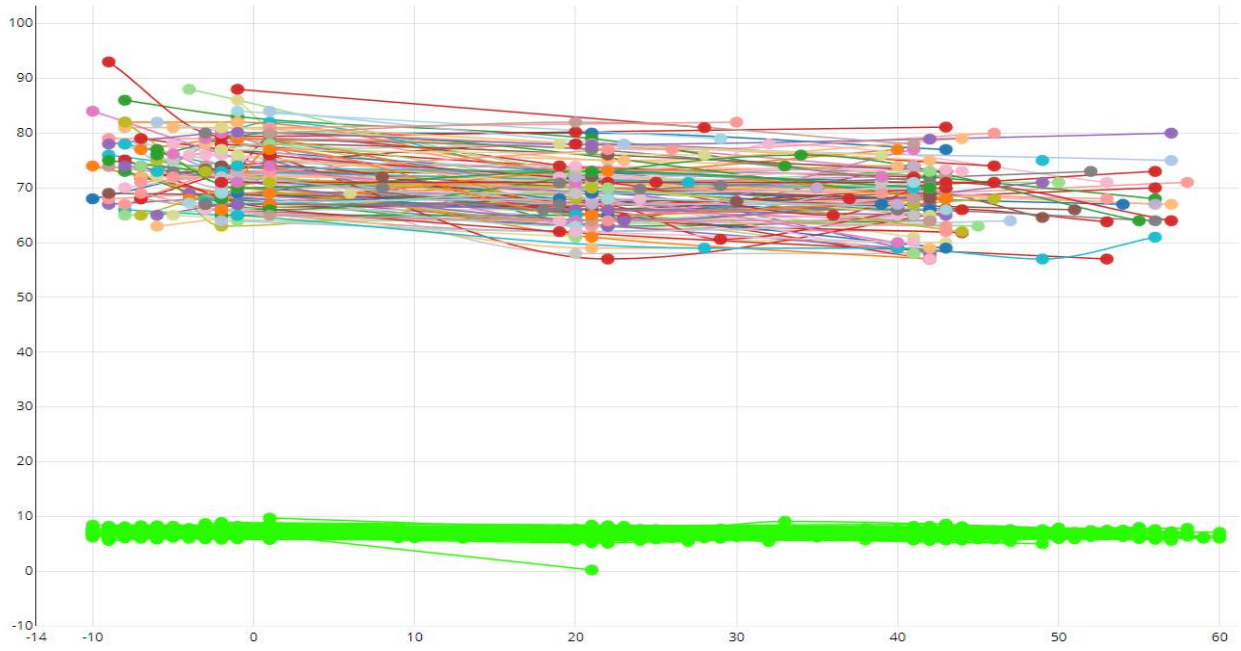

Figure 4. Example of outlying values: Patient data from Site D (green) manifest consistently outlying values, one order of magnitude different from the rest of the results.

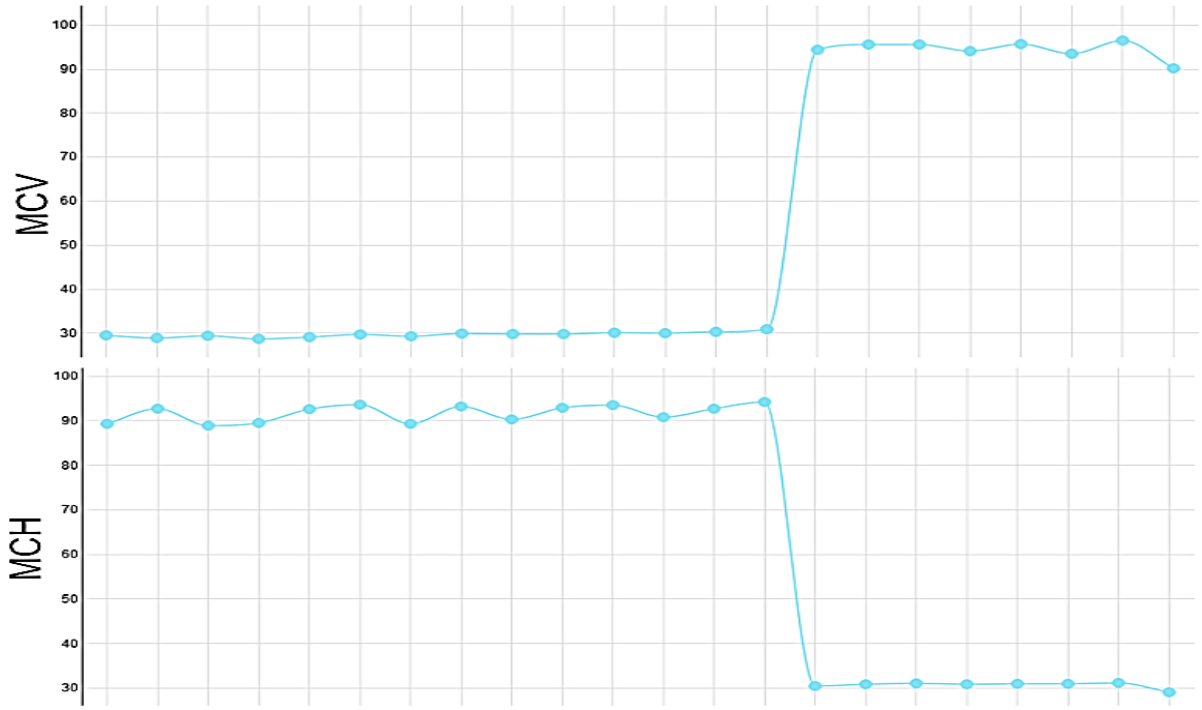

Figure 5. Example of data entry error in subject data – switched values of two simultaneously recorded results manifest in outlying values for both variables.

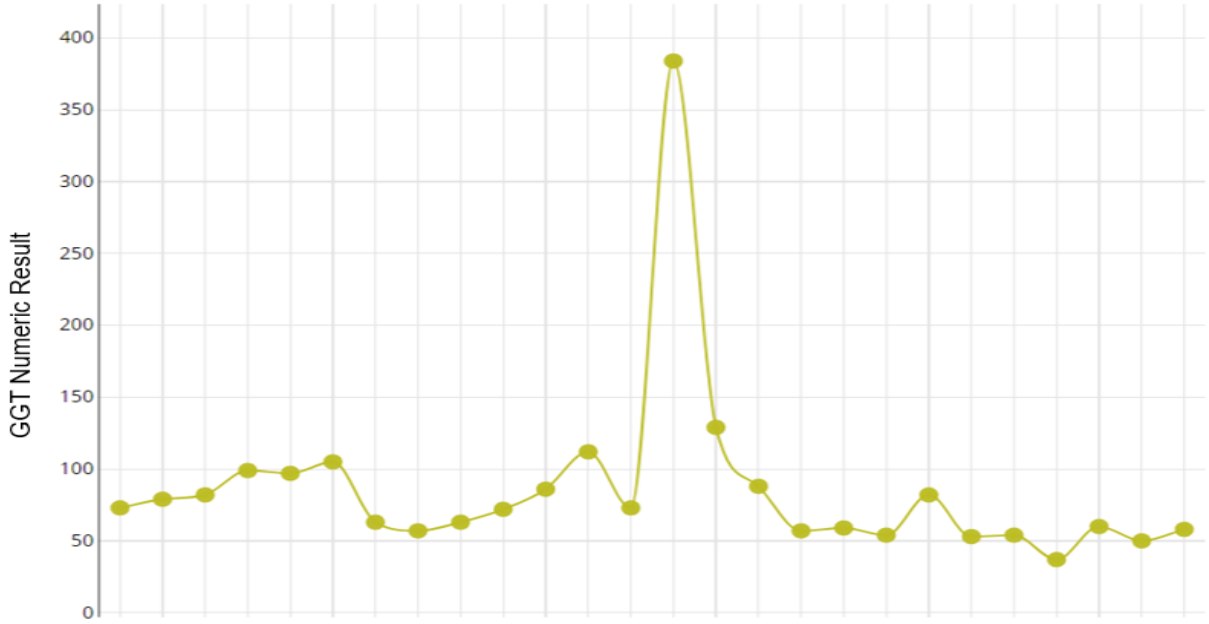

Figure 6. Example of an outlying value observed at individual subject level. In this case the reason might be a potential safety issue.

IMPLEMENTATION

Central statistical review assessment including score calculation results is periodically loaded into dashboard where all scores are connected with corresponding data visualizations. Central Statistical Monitor reviews the scores and variable distributions, and provides description of the pattern or signal observed in site or subject data. The observations and signal highlight recommendations are summarized in a review report. CSM findings are subsequently presented and discussed at meetings with study team and other relevant stakeholders, resulting in decisions on further actions.

CONCLUSION

The conclusion summarises your paper and ties together any loose ends. You can use the conclusion to make any final points such as recommendations predictions, or judgments. Finding anomalies and addressing quality issues in clinical trial data are effectively facilitated by applying unsupervised statistical methods integrated with graphical dashboard for visual data inspection. Incorporated in RBQM framework as part of Central Monitoring & Data Analytics activities, CSM helps to maintain data integrity during the course of the trial.

As CSM functions as a complement to other monitoring and data oversight activities, part of the role of Central Statistical Monitor is an active supply of information to and cooperation with other monitoring and data management roles. Regular interactions with Central Monitoring and Data Management roles as well as Medical Monitors are maintained as part of central statistical review process. Interactions with other monitoring roles are usually based on a specific action assigned to that role as a result of CSM observation.

CHALLENGES

Since data-driven monitoring is being successively being adopted to more trials, there are unfolding challenges in its implementation and conduct. Key aspects in this matter are: feasibility of processing large amount of data (e. g. in vaccine trials), delivery of data quality dashboards and visualisations, complexity of trial design, concentration of population cohorts (with characteristics specific for a location), and awareness of the meaning of CSM observations and associated risks.

FURTHER DEVELOPMENT

Development roadmap foresees CSM application earlier in the trial (with fewer data available) and directly on patient profile data, checks for potential duplicate patients and closing the gap between clinical and operational data reviews.

REFERENCES

Guidance for Industry; Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring (FDA, 2013); https://www.fda.gov/media/116754/download

Trotta L, Kabeya Y, Buyse M, et al. Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring. Clinical Trials. 2019;16(5):512-522. doi:10.1177/1740774519862564

Wilson B, Provencher T, Gough J, et al. Defining a Central Monitoring Capability: Sharing the Experience of TransCelerate BioPharma's Approach, Part 1; Therapeutic Innovation & Regulatory Science 2014; 48(5):529-535. doi:10.1177/2168479014546335

Monge ZA, Chen M, Lobanov VS, et al. A Hybrid Statistical-Machine Learning Approach to Anomaly Detection in Clinical Trial Data; 2020 APAC Symposium

ACKNOWLEDGMENTS

Fortrea – score calculation and analytics dashboard used for data visualization was provided via Xcellerate® Statistical Review tool;
PSI, ASA BIOP & EFSPI Centralized statistical monitoring and quality tolerance limits (CSM/QTL) joint scientific working group;
GSK Central Monitoring & Data Analytics Team.

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.) Your comments and questions are valued and encouraged. Contact the author at:

Wiktor Majewski
GSK
Rzymowskiego 53
02-697 Warsaw
Poland
Email: wiktor.w.majewski@gsk.com