



## Harnessing CSM to Drive Risk-Based Quality Management

By Laura Trotta, R&D Manager, CluePoints

### Abstract

Central Statistical Monitoring offers a data-driven approach to quality in clinical trials. Since a common protocol is used in a trial, all participating centres collect data on the same variables for the same visits. Every variable in the clinical database is potentially indicative of data quality, regardless of its meaning. Abnormal trends and patterns in the data can be detected by comparing the data across all centres. A Risk-Based approach to clinical research may include a central statistical assessment of data quality, which is crucial in the investigation of all possible risks and not just the obvious primary endpoints and safety data.

This paper will highlight, by example of a recent experiment, how a computer-intensive approach that generates large numbers of statistical tests is an effective way to check data quality and manage risk in multicentre clinical trials. The method offers a cost-effective complement to other data management and monitoring techniques.

### Introduction

In clinical trials, data quality is crucial. Any issues identified with the quality of the data may be linked to problems with how the trial was conducted. This could have possible safety implications for any patient subject involved in the trial. Data quality issues also have a detrimental effect on the approval process of the drug or device under investigation.

On-site monitoring of the clinical centres participating in a trial is one method to ensure the reliable assessment of the safety and efficacy of experimental treatments. Source-data verification, where the data collected in the case report form is compared with the original data in the patient's medical records is a time-consuming monitoring activity.<sup>1</sup> Source data verification takes up to half of the time of monitoring visits and imposes a high burden on site personnel.

The usefulness of this activity has been questioned and more cost-effective ways of conducting trials and checking data quality have been proposed.<sup>2</sup>

## Managing Risk

Managing risk can be done in a random, supervised or unsupervised way and the corresponding level of risk is either high, medium or low. For researchers managing risk in a clinical trial, a supervised approach is often adopted. The researchers can anticipate a potential issue or problem and therefore test for it and make sure it is avoided. This has its merit and value but what about the issues that can't be predicted? Implementing a Risk-Based approach into a clinical trial does not introduce new risks but enables the existing ones to be better managed. It all depends on the Risk Management Strategy (Random vs Supervised vs Unsupervised). Unsupervised Statistical Monitoring is aimed at detecting atypical data patterns in multicentre clinical trials, using the following principles:<sup>3</sup>

- a) data coming from the various centres participating in a trial should be largely similar, save for the random play of chance and systematic variations that occur in reality (e.g., due to patient ethnicity if centres are located in different regions of the world)
- b) a battery of standard statistical tests are applied to the patient data to compare the distribution of the data in one centre compared with all other centres
- c) Specific models such as mixed-effects models are used to allow for the natural variability between the centres;
- d) all tests that are relevant given the type of each variable (continuous, factor or date variable) are systematically applied to all patient-level data in a completely unsupervised manner

With recent findings demonstrating that the Food and Drug Administration (FDA) reject marketing approval for new drugs because of inconsistencies between trials, centres, or endpoints, or issues in study conduct<sup>4</sup>, a strategy for effectively managing risk is crucial. Technology is a catalyst to risk management, helping researchers to focus on what really matters (that may or may not have been anticipated). This is where the unsupervised 'leave no stone unturned' approach comes into its own. It is the ultimate safety net and insurance policy for a trial and all trials will benefit from the approach, both from a patient safety and ultimate success perspective.

## Case Study Example

A recent study was implemented to investigate the operating characteristics of unsupervised statistical central monitoring aimed at detecting atypical data in multicentre experiments. In this study, researchers used extensive simulations based on an actual large multicentre trial: The Stomach cancer Adjuvant Multi-Institutional group Trial (SAMIT, UMIN Clinical Trial Registry number C000000082), a large phase III trial conducted in Japan for patients with locally advanced, operable gastric cancer.<sup>5</sup>

In this study a statistical approach described previously for the detection of atypical data in multicentre experiments was used<sup>1</sup>. The approach is unsupervised, i.e. it requires no input from the user, and exhaustive, i.e. it uses all variables collected at the patient level. Central Statistical Monitoring using this approach has been shown to be effective at detecting atypical data that point to problems in actual studies.

In order to replicate a real-life situation, the trial data were contaminated by a group of researchers at Tokyo University for varying percentages of centres, percentages of patients modified within each centre, and numbers and types of modified variables. The unsupervised statistical monitoring software was run by a blinded team of researchers at CluePoints on the contaminated datasets, with the purpose of detecting the centres with contaminated data.

CluePoints was selected for the SAMIT trial because the Japanese investigators feared questions about the study conduct and data quality from Western journals. A full data quality assessment of the trial was performed using CluePoints Monitoring Platform. As it turns out, there were no data quality issues and the trial results were published in a high impact factor journal<sup>5</sup>. Therefore, data contamination was introduced in a clean trial.

The operating characteristics (sensitivity, specificity and Youden's J-index) were calculated for three detection methods used in the simulations:

1. One based on extreme P-values of individual statistical tests, adjusted for multiplicity to preserve the false discovery rate (FDR) in view of the number of statistical tests performed
2. Another using a summary of all P-values for a given centre, called the overall Data Inconsistency Score (DIS)
3. The last one combined both detection methods

## The Software Solution

CluePoints have established themselves in the market with active collaboration with FDA and the fact that Centralized Statistical Monitoring can be performed across all investigative site data, not just subject data is highly valuable. The CluePoints platform enables researchers to quickly discover potential outliers and erroneous data rather than waiting for issues to be identified during the submission process which can cause long and costly delays to pharmaceutical companies.

## Key Findings

The simulations conducted in the study are useful to assess the operating characteristics of an unsupervised statistical approach to the detection of data fabrication (or other types of systematic errors affecting the data) in clinical trials.

As the simulations were performed using data from an actual trial and contaminating these data over a wide range of parameters, they corresponded to both realistic situations of fraud and extreme situations unlikely to be seen in real life.

Results showed that the three methods were able to highlight data contamination anticipated in practice. Sensitivity increased in line with percentage of patients and variables contaminated. In all scenarios of contamination, the three methods showed a specificity better than 93 percent. The method based on the DIS and individual P-values adjusted for multiplicity generally had slightly higher sensitivity at the expense of a slightly lower specificity.

The method using extreme P-values requires proper adjustment to account for multiplicity, and this makes it unattractive as a general method to check data quality. The DIS, an overall score that summarizes inconsistencies across many variables and tests, has similar operating characteristics and may thus be preferred over an examination of many statistical tests taken in isolation. In addition, the DIS has the advantage of not only flagging centres with potential data issues, but also ranking the centres from the most atypical to the least atypical. Overall, the combined method using both the DIS and extreme P-values may be preferred because it has highest sensitivity, which implies it would err on the side of detecting more data issues, whether real or due to chance.

The analysts can choose the detection method and tune the threshold of statistical significance according to the type and size of the clinical trial being analysed. The SAMIT trial is representative of many multicentre clinical trials, with a few large centres and many small ones. P-values naturally take centre size into account, in such a way that in large centres data are considered atypical if they deviate less from the

average than in small centres. This is a desirable property of a statistical data quality assessment, but it may result in small data issues being left uncovered in centres of small size.

The results also demonstrated that the detection methods have the best operating characteristics when data are contaminated in fewer centres, for higher proportions of patients, and for larger numbers of variables, which is precisely the situation of a single centre having serious quality issues or committing fraud. This provides some reassurance that the worst cases of data fabrication will also generally be the easiest to detect.

The simulations show that continuous variables are more informative for the detection of atypical data than dichotomous or categorical variables, although this finding should be gauged considering the plausibility of the parameter values used in the simulations. No method had any power to detect contamination of a single categorical variable regardless of the number of patients or centres affected. In contrast, contamination of a continuous variable was detected if it affected most

patients in a single centre, particularly for variables repeatedly measured over time for all patients.

Detailed materials, methods, results and discussion can be found in the recently published paper: Detection of atypical data in multicentre clinical trials using unsupervised statistical monitoring.<sup>6</sup>

## Conclusion

The US Food and Drug Administration, the European Medicines Agency, and the International Committee on Harmonization all advocate using a risk-based approach to monitoring data quality in clinical trials. A central statistical assessment of data quality presents “opportunities for new monitoring approaches (e.g., centralized monitoring) that can improve the quality and efficiency of sponsor oversight of clinical investigations.”<sup>7</sup>

To address regulatory guidelines regarding a risk-based approach to monitoring, CluePoints has designed and developed a Central Statistical Monitoring software solution using advanced statistical tests to determine the quality, accuracy and integrity of clinical trial data, both during and after study conduct. These tests are based on non-predefined (unsupervised) criteria and, as such, are complementary to Key Risk Indicators. A whitepaper is available for further information and can be downloaded here: <https://cluepoints.com/white-papers-and-case-studies/>

The use of brute force (a computer-intensive approach that generates large numbers of statistical tests) to check data quality in multicentre clinical trials is at sharp variance with traditional methods based on careful reviews of individual patient data by trained personnel. The two approaches are best seen as complementary to each other. The simulation results presented in the case study suggest that an unsupervised approach using an overall data inconsistency score has good operating characteristics to detect centres with atypical data that are likely to occur in practice.

This approach is advantageous in terms of objectivity, reproducibility, and cost. It can be used to inform and expedite site reviews by trained personnel in a Risk-Based monitoring framework. In the SAMIT trial, for instance, the findings of Central Statistical Monitoring led to site audits that confirmed minor issues without any potential impact on the trial outcome. The elimination of source data verification, except when triggered by atypical data, would free up a considerable amount of human resources that could be spent on more important risks in clinical trials, particularly those related to patient safety and other key aspects of trial conduct.

## References

1. Morrison BW, Cochran CJ, White JG et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. *Clinical Trials* 2011; 8(3): 342–349.
2. Sheetz N, Wilson B, Benedict J et al. Evaluating source data verification as a quality control measure in clinical trials. *Therapeutic Innovation & Regulatory Science* 2014; 48(6): 671–680.
3. Venet D, Doffagne E, Burzykowski T et al. A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials* 2012; 9(6): 705–713.
4. Sacks LV, Shamsuddin HH, Yasinskaya YI et al. Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000–2012. *JAMA* 2014; 311(4): 378–384.
5. Tsuburaya A, Yoshida K, Kobayashi M et al. Sequential paclitaxel followed by tegafur and uracil (UFT) or S-1 versus UFT or S-1 monotherapy as adjuvant chemotherapy for T4a/b gastric cancer (SAMIT): a phase 3 factorial randomised controlled trial. *The Lancet Oncology* 2014; 15(8): 886–893.
6. Trotta, Laura & Kabeya, Yuusuke & Buyse, Marc & Doffagne, Erik & Venet, David & Desmet, Lieven & Burzykowski, Tomasz & Tsuburaya, Akira & Yoshida, Kazuhiro & Miyashita, Yumi & Morita, Satoshi & Sakamoto, Junichi & Praveen, Paurush & Oba, Koji. (2019). Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring. *Clinical Trials*. 174077451986256. 10.1177/1740774519862564.
7. US Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Oversight of Clinical Investigations
8. A Risk-Based Approach to Monitoring. <https://www.fda.gov/downloads/Drugs/Guidances/UCM269919.pdf> (2013, Accessed 16 July 2017)

