Paper AS05

How Central Statistical Monitoring Drives a Risk-Based Approach to Clinical Trial Monitoring

Charles Warne, Cytel, Singapore

ABSTRACT

By performing analyses at regular intervals on accumulating clinical trial data, Central Statistical Monitoring (CSM) provides Sponsors with valuable and timely information about which sites may be experiencing data, conduct or safety issues. This can inform a risk-based allocation of site monitoring resources, enabling the most impactful corrective actions to be taken before study integrity is compromised. As a result, CSM can reduce the percentage of Source Document Verification (SDV) required and thereby save monitoring costs, while giving Sponsor's greater visibility and assurance of the integrity of their data before submitting it to regulatory authorities (who routinely perform the same types of analyses). This paper provides an overview of CSM, what it entails, when it adds value, how it embodies Risk-Based Quality Management (RBQM) principles, and how it drives a risk-based approach to clinical trial monitoring, including practical examples from real case studies.

INTRODUCTION

Central Statistical Monitoring can be defined as the statistical detection of anomalies in accumulating clinical trial data to identify groups (i.e. sites, patients, countries) that are performing differently to the rest, where the reasons for such differences may relate to issues in data quality, site conduct or safety. The origins of CSM can be traced back to work by Buyse and colleagues (Buyse, 1999) who proposed statistical methods for fraud detection in clinical trials. Although their paper focused on the issue of fraud, it provided a statistical framework for detecting a spectrum of data issues, i.e. ranging from fraud (intention to cheat) -> tampering (deliberate, but without intention to cheat) -> sloppiness (incorrect reporting due to poor awareness) -> errors (unintentional technical problems) (Buyse, PHUSE). While fraud at a single site can tarnish the credibility of a study, it is relatively rare and does not usually impact the scientific conclusions. The other types of data issues (i.e. sloppiness/errors/tampering) are more common, making up the vast majority of CSM findings. Their impact can be substantial if they affect critical endpoints or their prevalence calls into question the overall data integrity and undermines confidence in study results.

BENEFITS OF CSM

Central Statistical Monitoring can detect data issues that would be missed by a monitoring strategy that only relies on SDV, or a data cleaning strategy that only relies on standard Data Management (DM) checks. This is because SDV and standard DM checks are not designed to detect anomalies within or between patients, at a site or study level, or in the multivariable space. By flagging 'at risk' sites or patients, CSM supports a risk-based approach to monitoring and is a powerful tool for data quality oversight.

The benefits of CSM can be grouped into two broad categories:

(1) Smarter resource allocation, leading to:
- Better return on investment (on-site monitoring targeted to areas of highest impact)
- Cost savings (due to reduced SDV, which typically accounts for ~15% of a clinical trial's budget)

(2) Higher data quality, leading to:
- Reduced timelines (faster enrollment and database lock)
- Reduced regulatory risk
- Cost savings (less re-work)

REGULATORY EVOLUTION TOWARDS RISK-BASED & CENTRALISED MONITORING

Guidance documents from health authorities have progressively endorsed a risk-based approach to monitoring, including the key role of centralised monitoring:
- FDA (2004) introduced risk-based approaches and Quality by Design (QbD) principles for drug development.
- FDA (2013) and EMA (2013) published guidance on risk-based approaches to monitoring and quality management in clinical trials.

- ICH E6(R2) (2016) endorsed a risk-based approach to the extent and nature of monitoring, formalised the concept of centralised monitoring, and recommended that Sponsor's consider a combination of central and on-site monitoring for their trials, including the possibility of central monitoring alone.
- MHRA (2022) published guidance on oversight and monitoring activities, and described the role of central and statistical monitoring with examples.
- The recently published ICH E6(R3) (2025) integrated concepts of risk-proportionality and Quality by Design (QbD) into Good Clinical Practice (GCP) principles, and added a dedicated section (*3.11.4.2 Centralised Monitoring*) to reinforce that central monitoring can identify systemic or site-specific issues and inform targeted site monitoring.

COMPONENTS OF CENTRAL MONITORING

Central monitoring typically encompasses three types of analyses, each summarised below.

(1) Key Risk Indicators (KRIs):
- Leading metrics for detecting risks to data quality, conduct, safety
- Simple summary measures, i.e. rates of serious/non-serious AEs, protocol deviations, missed assessments, early terminations, screen failures, auto-queries, eCRF visit to entry cycle time.
- Analysed at site-level
- 10-20 per study

(2) Quality Tolerance Limits (QTLs):
- A special subset of KRIs focusing on critical risks
- Analysed at study-level
- 3-5 per study
- Thresholds defined from comparable studies

(3) Central Statistical Monitoring (CSM):
- Data anomaly detection
- Advanced statistical methods applied to high-volume data
- Flags potential issues in data quality, site conduct or safety that warrant attention
- Usually analysed at site-level, but can also be subject- or country-level

CSM: WHEN TO APPLY, HOW OFTEN, DATA TYPES AND ANALYSES

ELIGIBILITY FOR CSM

Certain criteria need to be met for CSM analyses to be viable. For site-level comparisons, the study should have a sample size of at least ~30 patients and approximately four sites enrolling at least five patients. This is the bare minimum, and a higher number of sites and patients is recommended for anomaly detection to be more powerful. For smaller studies where these criteria are not achievable, patient-level comparisons may be considered if there are sufficient data collected on each patient. Study duration is also important; not only because a longer time on study generates a richer set of repeated measurements over time, but also because longer studies allow for periodic CSM runs and enough time for early corrective and preventative actions to be beneficial. For this reason, CSM can be considered for studies of at least six months duration.

HOW OFTEN TO APPLY CSM ON A STUDY

The timing of analyses can vary depending on the rate of data collection, the risk profile of the study and the degree to which central monitoring is integrated into the overall monitoring plan. Typically, a higher frequency is warranted during enrolment/treatment (i.e. ~3 monthly), reducing to longer intervals at follow-up (i.e. ~6 monthly). For a leaner approach, Sponsors may choose to perform only one CSM analysis to verify the integrity of the data before a critical study milestone (i.e. Interim Analysis, Dry-run, Database Lock). Finally, the timing can also be risk-based, with frequency adapted to observed findings.

TYPES OF DATA

A range of clinical and operational data can be analysed using CSM, including eCRF/EDC (raw), SDTM/ADaM (derived), external data (labs, biomarkers, eDiary, PRO), IxRS/IRT, CTMS, queries, and audit trail data. There is no limitation to the types of data that can be interrogated provided they are available in machine-readable format, but it helps to consider which provide the greatest return (towards ensuring study integrity) for the calibration effort needed. For example, consider a study where biomarker assessments are only performed once on a subset of patients. It may

not be worthwhile including these data in CSM analyses if there are many variables to calibrate but an insufficient number of patients for site-based comparisons. In contrast, although a significant amount of study-specific calibration is required to ingest raw EDC data compared to SDTM/ADaM, it may be justified because raw data extracts are usually available much earlier than SDTM/ADaM and therefore findings can be shared with the study team sooner.

TYPES OF ANALYSES

Early analytical approaches to CSM were 'supervised', employing custom tests on a reduced set of variables, endpoints and KRIs. While computationally simpler, they were less sensitive at detecting data issues. More recent methods apply an 'unsupervised' brute force computer-intensive approach, performing an exhaustive set of generic tests to a larger set of variables. This yields higher sensitivity to detect data issues and is now the industry gold standard.

The tests performed by CSM can be defined according to whether they involve a single variable or multiple variables. Single variable tests compare one site to the rest using statistics appropriate to the type of variable, i.e. the mean and standard deviation (within and between-patients) for continuous variables, the binary proportion for k-1 levels of a k-level categorical variable, or the rate for count variables. The digit preference distribution of a continuous variable can be tested by comparing how one site compares to the rest on the proportion of all observations with a 0, 1, 2, 3, 4, 5, 6, 7, 8 or 9 as their first/last digit. Temporal variables are a special case; anomalies may be detected by converting their date/time components into numerical quantities for which the mean and standard deviation can be compared across sites, while calendar-based methods compare the proportion of visits occurring on a weekend or public holiday. Multiple-variable tests detect outliers by considering if the distribution of a set of variables within a patient, or between patients at a site, is statistically different to the rest. This includes looking for multivariate outliers (i.e. results far from the multivariate mean), multivariate inliers (i.e. results close to the multivariate mean), unexpected correlations between variables (within and between patients), and 'professional patients' based on an unlikely similarity of data across different patient IDs.

After all tests are performed, the resulting p-values for a site can be negative log-transformed to obtain a risk score for each variable and test, where a higher risk score signifies a higher degree of difference between that site and the rest. A site's risk score can be summed across all variables to provide a summary risk index for each test, and finally the risk indexes can be summed across all tests to obtain a global risk index for each site. These hierarchies of summation support the review of results by providing an overall bird's eye view comparing one site to the rest, while also the ability to drill-down to the individual variable level for each test.

INSIGHTS FROM CSM DRIVE RISK-BASED MONITORING AND SUPPORT RISK-BASED QUALITY MANAGEMENT

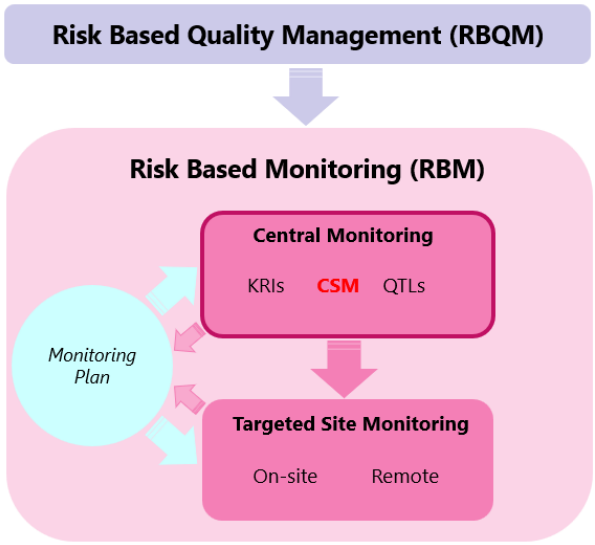
Risk-Based Monitoring (RBM) plays a crucial role within the Risk-Based Quality Management (RBQM) ecosystem of a clinical trial. In parallel with regulatory developments over the past 15 years, groups such as TransCelerate and the Clinical Trials Transformation Initiative (CTTI) have taken a lead role in advancing the concepts, methodologies and tools for Sponsors to implement RBM into their trials.

As highlighted by TransCelerate (www.transceleratebiopharmainc.com), quality risk management principles are embedded into the RBM philosophy:

- Building Quality by Design (QbD) into trials
- Critical Processes and Data ('critical to quality factors')
- Early and ongoing Risk Assessment
- Use of Risk Indicators and Thresholds
- Ongoing adjustment of monitoring activities based issues and risks that may arise throughout the study

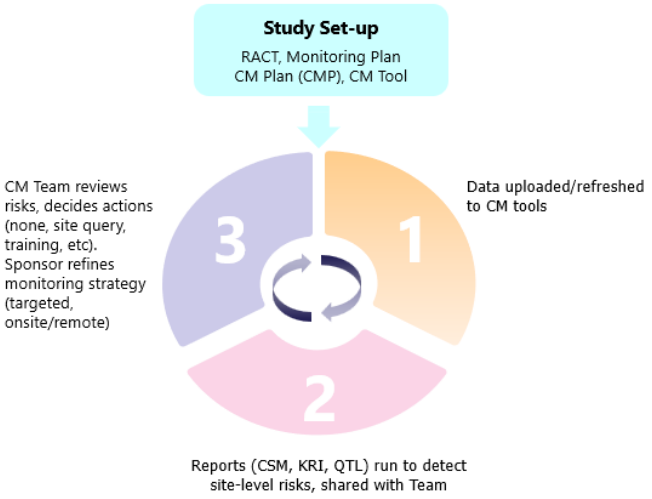
The Clinical Trials Transformation Initiative defines quality in clinical trials as 'the absence of errors that matter to decision making – that is, errors which have a meaningful impact on the safety of trial participants or the credibility of

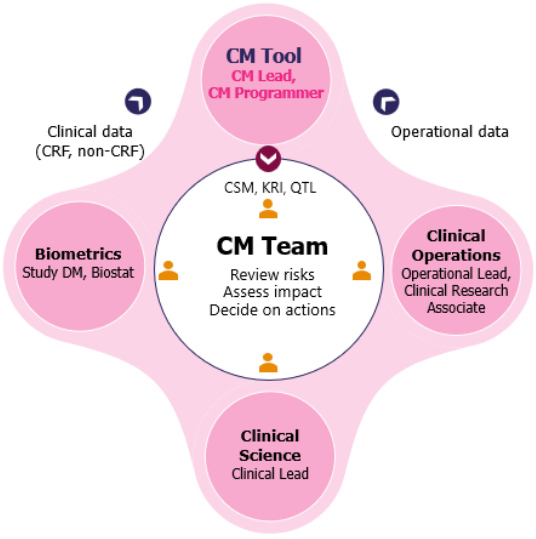
the results'. By performing regular analyses on high volumes of clinical and operational data, CSM can efficiently and objectively 'shine a light' on data anomalies to expose higher risk sites that may warrant further scrutiny. Coupling this with metrics from KRIs and QTLs, central monitoring provides valuable insights about where to target site monitoring efforts for the greatest return. This requires a flexible monitoring plan to adjust the allocation of on-site monitoring resources to new information, keeping in mind that the overall goal is to safeguard critical to quality factors by addressing the 'errors that matter to decision making'.



OPERATING MODEL & MULTI-DISCIPLINARY TEAM STRUCTURE

The diagram (below left) outlines the CSM process for a study, noting that other components of central monitoring (KRIs and QTLs) may be analysed and reviewed in parallel if they are in scope and share the same reporting frequency.





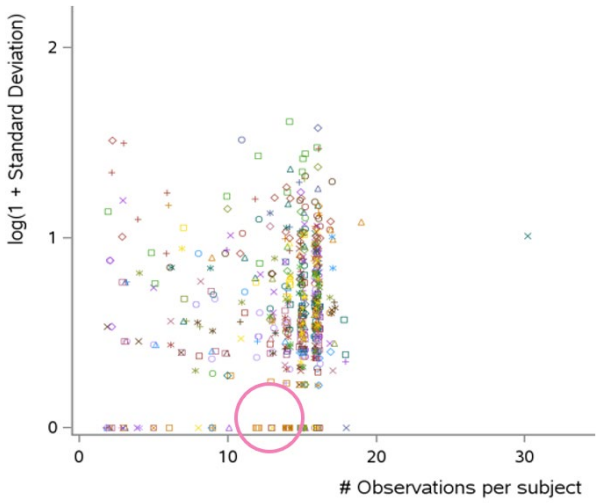
CM = Central Monitoring, RACT = Risk Assessment Categorisation Tool

Prior to enrolling the first patient, a Central Monitoring Plan (CMP) is developed based on the study protocol and CRF to specify the scope and timing of each analysis (CSM, KRIs, QTLs as needed) and the classification of data types. The central monitoring tool is then calibrated to meet individual study specifications for each data set. After each analysis run, the results are reviewed and key findings are compiled in a report. This report is shared with the central monitoring team and discussed at a results review meeting, where it is decided what follow-up actions are needed to address confirmed issues and mitigate emerging risks. The goal of this meeting is also to review the status of findings from previous reporting events, track the performance of sites over time to assess the impact of any corrective actions, and inform a risk-based approach to site monitoring. The central monitoring team (see diagram above right) brings together a variety of cross-functional expertise from groups including biometrics, clinical operations and clinical

science. This reflects the diversity of perspectives and skill sets needed to review the data, understand its limitations, interpret results, and decide on follow-up actions to ensure the integrity of critical to quality factors.

CASE STUDY #1: DATA PROPAGATION

In this example, CSM analyses of a global multicenter trial revealed outlying sites based on the within-patient SD for respiratory rate. At these sites, patients' respiratory rate results were replicated across all visits (corresponding to a within-patient SD=0). The graph below plots the log(1+SD) against the number of observations per subject, with markers shown separately for each subject and colour coded by site. The points along the bottom of the graph correspond to a within-patient SD=0.



At one site (#23), 8 out of 12 patients had respiratory rate results replicated across all visits (below left). An example is shown for one patient (ID=23001, below right) who had a respiratory rate of 18 breaths per minute recorded across all 14 visits.

Site ID	Subject ID	# Obs. per subject	Standard Deviation	log(1 + Standard Deviation)
23	23001	14	0.0000	0
23	23002	15	0.2582	0.2296812
23	23003	15	0.2582	0.2296812
23	23004	3	0.0000	0
23	23005	15	0.0000	0
23	23006	14	0.0000	0
23	23007	10	0.3162	0.2747699
23	23008	15	0.0000	0
23	23009	14	0.0000	0
23	23010	14	0.0000	0
23	23011	13	0.2774	0.2447877
23	23012	15	0.0000	0

Site ID	Subject ID	Visit ID	Result	Units
23	23001	10	18	breaths/min
23	23001	20	18	breaths/min
23	23001	30	18	breaths/min
23	23001	40	18	breaths/min
23	23001	50	18	breaths/min
23	23001	60	18	breaths/min
23	23001	70	18	breaths/min
23	23001	80	18	breaths/min
23	23001	90	18	breaths/min
23	23001	100	18	breaths/min
23	23001	110	18	breaths/min
23	23001	120	18	breaths/min
23	23001	130	18	breaths/min
23	23001	180	18	breaths/min

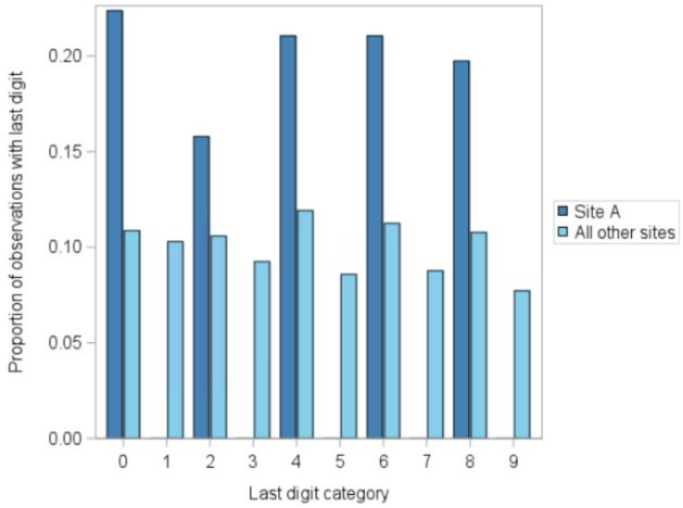
Note: Site and subject IDs are dummy coded

This lack of variation within patients is clinically implausible, and is an example of 'propagation', where the baseline value is incorrectly carried forward to all other visits. This could be due to inadequate data collection practices at site, or downstream data processing errors in the EDC system.

CASE STUDY #2: LAST DIGIT PREFERENCE DISTRIBUTION

A last digit preference CSM analysis was performed on a multicenter trial, revealing an outlying distribution for a lab parameter at one site. At Site A, a total of 76 observations were recorded for the lab parameter across 5 patients, with the raw data (left) and percentage of measurements with each last digit (right) shown below:

60	72	60	54	84	102	138	84
60	66	66	66	90	96	90	96
60	66	72	60	96	90	90	36
60	60	54	48	102	96	78	42
90	60	48	48	102	66	84	48
78	66	54	54	96	60	84	48
72	66	60	54	120	66	78	
66	54	54	54	108	84	78	
72	48	48	72	96	72	84	
90	72	48	78	102	84	84	



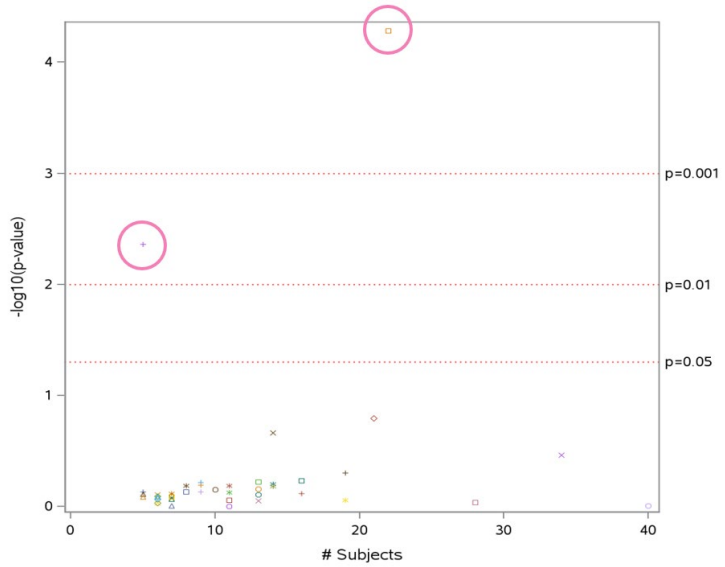
The plot resembles a classic 'combed distribution' (dark blue bars) for Site A, where remarkably the last digit is an even number for all 76 measurements of this lab parameter. This contrasts to the relatively uniform last digit distribution seen for the remaining sites (light blue bars). The majority of p-values comparing the proportion of measurements with each last digit at Site A to the rest of sites are statistically significant at the $p<0.01$ (n=4) or $p<0.05$ (n=4) level:

last_ digit	site_ count	site_ proportion	overall_ proportion	p_value
0	17	0.224	0.109	0.0058
1	0	0.000	0.103	0.0044
2	12	0.158	0.106	0.2055
3	0	0.000	0.092	0.0659
4	16	0.211	0.119	0.0323
5	0	0.000	0.086	0.0423
6	16	0.211	0.112	0.0288
7	0	0.000	0.088	0.0019
8	15	0.197	0.108	0.0295
9	0	0.000	0.077	0.0064

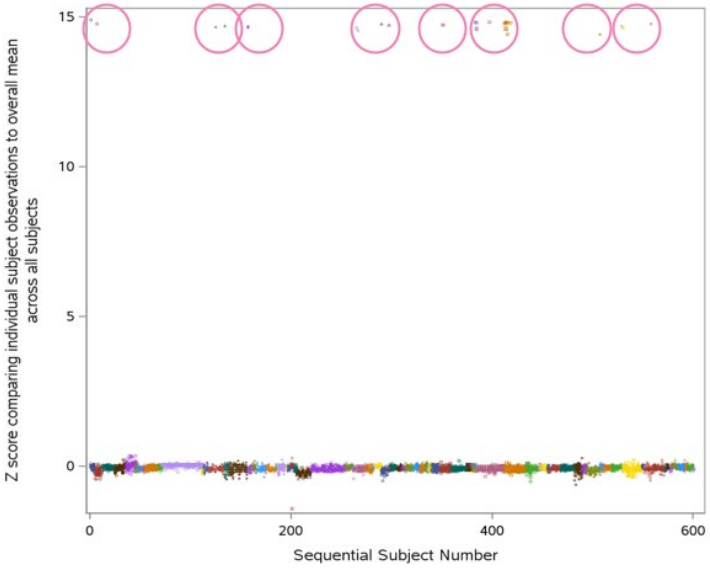
This finding is unusual and warrants further investigation to determine its root cause. It may be due to unintentional equipment miscalibration, conversion issues, tampering, or even fraud.

CASE STUDY #3: INCORRECT UNITS

In this example, a multicenter trial was subjected to CSM analyses which revealed two sites with extreme mean temperatures (40.2 ºC, 44.9 ºC) compared to the rest (36.9 ºC). The plot below shows the -log(p-value) comparing mean temperatures at each site to the remaining sites against the number of subjects per site, with the two outlying sites highlighted:



Upon further investigation it was found the issue was more systemic, affecting a larger number of sites. This can be seen when plotting the Z scores comparing individual subject means to the overall mean against the site/subject number (below left), revealing several patients at different sites with extremely high temperatures. As shown (below right), based on the magnitudes observed it appears some temperatures were recorded in Fahrenheit but units were fixed to Celsius in the EDC. Ideally this should be addressed by retraining sites to enter the data in Celsius. Alternatively, the EDC could be updated to allow units of Fahrenheit, or as a last resort the outlying results could be converted at the analysis stage.



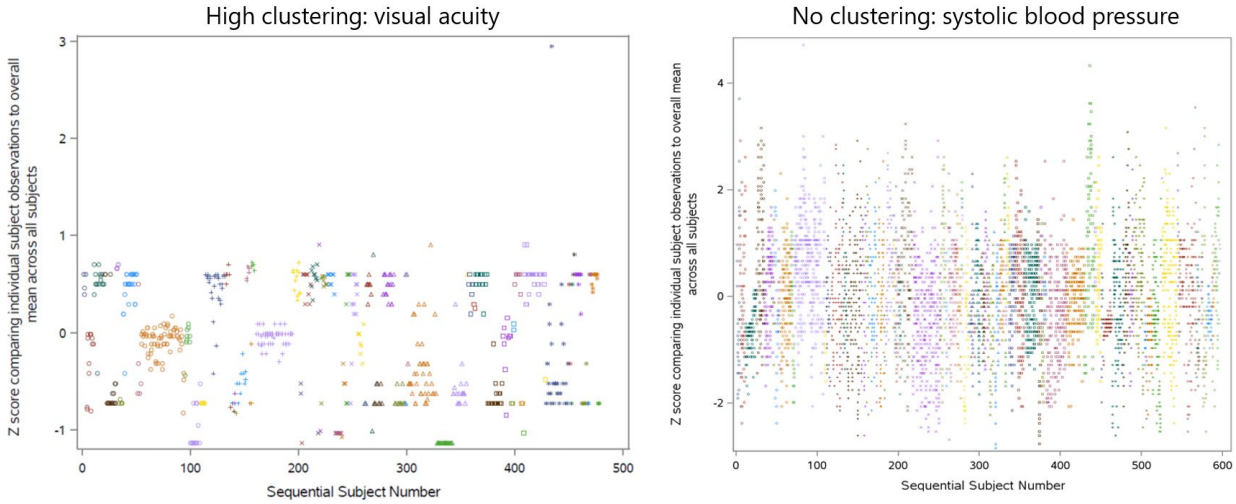
Site ID	Sequential Subject Number	Visit ID	Result	Units
140	157	10	97.6	C
140	158	10	36.2	C
140	159	10	36.2	C
140	160	10	36.4	C
140	161	10	36.2	C
381	412	10	98.1	C
381	413	10	97.6	C
381	414	10	36.8	C
381	415	10	96.6	C
381	416	10	36.7	C
381	417	10	37	C
381	418	10	98.2	C
381	419	10	36.7	C
381	420	10	36.1	C

Note: Site and subject IDs are dummy coded

CASE STUDY #4: SITE-LEVEL CLUSTERING

In this final example, CSM analyses of a multi-center trial revealed a high degree of site-level clustering for a visual acuity endpoint (Intraclass Correlation Coefficient ~ 0.9). This differs from previous case studies, in that it's a study-level finding rather than a case of one site performing differently to the rest. The root cause for such a high degree of clustering may be that different testing methods or equipment were used across sites, leading to a lack of standardisation in how visual acuity was measured. It may warrant additional training of investigators or a review of the testing protocol. At a minimum, the clustering should be accounted for in the final analysis.

Visualisations are an essential part of CSM analyses. In this example, they were able to powerfully highlight the degree of site-level clustering by plotting the Z scores comparing individual observations to the overall mean against the site/patient number ('Sequential Subject Number'). There is a striking difference that can be seen in the distribution of data for visual acuity (below left) compared to a typical variable with negligible site-level clustering (systolic blood pressure, below right).



This reinforces the value of a broad-based approach to CSM analyses, where additional statistics and graphical techniques can highlight anomalies that may otherwise go unnoticed when, for example, the focus is limited to computing p-values from site-level comparisons of means and SDs.

CHALLENGES TO IMPLEMENTATION

There are several additional points to consider when implementing CSM.

The first is that a sufficient volume of data are needed before the initial CSM run, yet the number of subjects and assessments at the beginning will naturally be low and site initiation tends to be staggered. This requires a balancing act to gauge when enough data have accumulated without waiting too long for issues to be detected, because preventive and corrective actions will have their greatest impact when implemented early. A good awareness of the data flow and enrolment rates per site is needed to optimize the timing of the first CSM run and the cadence of subsequent reporting events.

Data may not be clean before the early CSM runs, which can add noise to the signal. However, it is expected that some of the data issues will be due to unclean data, and if CSM can detect and resolve these earlier it will benefit the study overall. It helps to work closely with the study DM to understand when cleaning efforts are planned, and indeed whether CSM runs should be timed to inform data cleaning.

For blinded trials, it is important to evaluate upfront if any data could potentially unblind a patient's treatment. Although CSM reports would not include variables for the actual or randomised treatments, there is a chance that displaying other types of data in patient-level listings may be unblinding to the trained eye (i.e. pharmacokinetic, biomarkers, laboratory data). Decisions on whether to exclude such potentially unblinding data should be taken with the study team and documented in the Central Monitoring Plan before the first CSM analysis, to mitigate the risk of accidental unblinding.

By comparing one site to the rest, anomaly detection is a relative exercise which fundamentally assumes most sites are performing well. If this assumption does not hold, site-based analyses would not be expected to identify sites

worthy of attention and may not add value. Instead, external QTL/KRI benchmarks may be more useful to understand how the study is performing overall.

Finally, valid reasons may exist for why one site is being flagged as an outlier, such as age, race, socio-economic status or cultural differences. These should always be borne in mind to ensure the right variables are selected for CSM and results are interpreted in context.

CONCLUSION

Central Statistical Monitoring drives a risk-based approach to monitoring and quality management for clinical trials by helping study teams identify where the risks are, so they can target on-site monitoring resources effectively.

This is achieved:

- Using statistical methods that detect anomalies more effectively than manual review methods such as SDV/SDR
- Based on the totality of data collected (where high volume leads to better performance)
- Using objective measures of comparison, applied in a standardised way across different types of clinical trials and disease areas
- In an efficient and reproducible manner (via statistical programming code)

REFERENCES

Buyse M., et al. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. Statistics in Medicine 1999; 18: 3435-3451.

Buyse M. Role of Biostatistics in Fraud Detection, PHUSE presentation (slides).

Clinical Trials Transformation Initiative: https://ctti-clinicaltrials.org/about/ctti-projects/quality-by-design/

European Medicines Agency, Reflection paper on risk based quality management in clinical trials, 2013.

United States Food and Drug Administration, Pharmaceutical CGMPs for the 21st Century – A risk-based Approach (Final Report). 2004.

United States Food and Drug Administration, Guidance for Industry: Oversight of Clinical Investigations – A Risk-based Approach to Monitoring, 2013.

George S., Buyse M. Data fraud in clinical trials. Clinical Trial Perspective 2015; 5(2), 161-173.

International Council for Harmonisation, Guideline for Good Clinical Practice E6(R2), 2016.

International Council for Harmonisation, Guideline for Good Clinical Practice E6(R3), 2025.

Medicines & Healthcare products Regulatory Agency, Guidance: Oversight and Monitoring Activities, 2022 (https://www.gov.uk/government/publications/oversight-and-monitoring-of-investigational-medical-product-trials/oversight-and-monitoring-activities)

TransCelerate Biopharma Inc: https://www.transceleratebiopharmainc.com/rbminteractiveguide/how-does-clinical-trial-site-monitoring-work-under-a-risk-based-monitoring-approach/the-transcelerate-model/#:~:text=Building,the%20risks%20may%20be%20managed.

Venet D., et al. A statistical approach to central monitoring of data quality in clinical trials. Clinical Trials 2012; 9(6), 705-713.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Author Name: Charles Warne
Company: Cytel
Email: charles.warne@cytel.com
Website: www.linkedin.com/in/charles-warne

Brand and product names are trademarks of their respective companies.