

# Central statistical monitoring in clinical trial management: a scoping review

Total number of words: 4 234

Maciej Fronc<sup>1,2\*</sup>: <http://orcid.org/0000-0002-4874-950X> (corresponding author)

Michał Jakubczyk<sup>1</sup>: <http://orcid.org/0000-0002-0006-6769>

Sharon B. Love<sup>3</sup>: <https://orcid.org/0000-0002-6695-5390>

Susan Talbot<sup>4</sup>: <https://orcid.org/0000-0002-7713-2226>

Timothy Rolfe<sup>5</sup>: <http://orcid.org/0000-0001-8798-840X>

<sup>1</sup> Decision Analysis and Support Unit, Institute of Econometrics, SGH Warsaw School of Economics, Warsaw, Poland

<sup>2</sup> Central Monitoring and Data Analytics, GSK, Warsaw, Poland

<sup>3</sup> MRC Clinical Trials Unit at UCL, UCL, London, UK

<sup>4</sup> Biostatistics, Amgen, Thousand Oaks, California, USA

<sup>5</sup> Central Monitoring and Data Analytics, GSK, Brentford, Middlesex, UK

Correspondence details for the \*corresponding author:

Decision Analysis and Support Unit, Institute of Econometrics,

SGH Warsaw School of Economics,

Al. Niepodległości 162, 02-554 Warsaw, Poland

e-mail: [mf85106@doktorant.sgh.waw.pl](mailto:mf85106@doktorant.sgh.waw.pl)

## **Abstract**

**Background:** Clinical trials handle a huge amount of data which can be utilised during the trial to improve the ongoing study conduct. It is suggested by regulators to implement the remote approach to evaluate clinical trials by analysing collected data. Central statistical monitoring helps to achieve that by employing quantitative methods, the results of which are a basis for decision-making on quality issues.

**Methods:** This article presents a scoping review which is based on a systematic and iterative approach to identify and synthesise literature on central statistical monitoring methodology. In particular, we investigated the decision-making processes (with emphasis on quality issues) of central statistical monitoring methodology and its place in the clinical trial workflow. We reviewed papers published over the last ten years in two databases (Scopus and Web of Science) with a focus on data mining algorithms of central statistical monitoring and its benefit to the quality of trials.

**Results:** As a result, 24 scientific papers were selected for this review, and they consider central statistical monitoring at two levels. Firstly, the perspective of the central statistical monitoring process and its location in the study conduct in terms of quality issues. Secondly, central statistical monitoring methods categorised into practices applied in the industry, and innovative methods in development. The established methods are discussed through the prism of categories of their usage. In turn, the innovations refer to either research on new methods or extensions to existing ones.

**Discussion:** Our review suggests directions for further research into central statistical monitoring methodology – including increased application of multivariate analysis and using advanced distance metrics – and guidance on how central statistical monitoring operates in response to regulators' requirements.

Keywords: clinical trials, central statistical monitoring, central monitoring, risk-based monitoring, risk-based quality management, quality by design, data-driven approach

## Background

According to the U.S. Food and Drug Administration (FDA), centralised monitoring is defined as ‘a remote evaluation carried out by sponsor personnel or representatives (e.g., clinical monitors, data management personnel, or statisticians) at a location other than the sites at which the clinical investigation is being conducted’.<sup>1</sup> The European Medicine Agency (EMA) claims there is a potential to improve central monitoring to central statistical monitoring by using statistical methodology to monitor the quality of the trial conduct and data.<sup>2</sup> Central statistical monitoring is the most efficient way to ensure patient safety, trial integrity and data quality in multicentre clinical trials,<sup>3-5</sup> and its use is recommended by both FDA and EMA, and in the ICH-E6(R2) Good Clinical Practice guidelines.<sup>1,2,6</sup> Early implementation of central monitoring models focussed primarily on key risk indicators as a simple, implementable solution for identifying site to site variation.<sup>7,8</sup> However, a more holistic approach to the data quality assessment via using complex statistical analytics enriches simple univariate assessment of key risk indicators and provides data-driven solutions for detection of systemic issues, data irregularities and potential fraud. Information obtained from central statistical monitoring analyses is a roadmap of decision-making processes.<sup>9,10</sup>

Although it might seem reasonable to assume that the incidence of misconduct or fraud in science in general and in clinical trials, in particular, is low; the true incidence is difficult to estimate.<sup>5</sup> This uncertainty recently led Smith to conclude that the problem of fraud in medical research ‘is huge’.<sup>11</sup> This uncertainty highlights the need for additional research in the area of trial quality, with a specific focus on fraud detection methodology.

Kirkwood et al. brought together the thinking on central statistical monitoring and published R-programs for others to use.<sup>12</sup> They classified individual participant-level monitoring and site-level monitoring that was required and gave methods to look for recording and entry

errors, procedural errors and fraud. They envisioned a time in the future when these checks would be automated and routinely carried out. Almost a decade later, trial sponsors have interpreted which checks to run and how and when to run them and many have extended the methodologies. However, limited information is available on the methods used and the results of any checks are not routinely reported.

This review aims to answer the question of how central statistical monitoring is conducted in the process of the clinical trials with an emphasis on quality issues, and what kind of quantitative methods are applied already or are being developed. This approach provides practical knowledge about central statistical monitoring in terms of technical solutions and how it is used.

In order to answer these questions, the review of central statistical monitoring literature was conducted according to the approach on a scoping review recommended by Arksey and O'Malley and supported by the PRISMA guideline by mapping key concepts underpinning the chosen research area settled in a broader theoretical framework.<sup>13-15</sup> The chosen approach helps to identify in central statistical monitoring research gaps and directions, investigate current practices, and also their barriers.

## **Methods**

### ***Literature search and method extraction***

Two large citation electronic databases were searched – Scopus and Web of Science.

In the search, we used the following keywords which define the research axis of the review (Table 1). The query was developed in agreement between all authors of this paper.

<Table 1>

The search was conducted on papers published in English between January 2012 and April 2022 to give a 10 year span. We excluded from the search books and book chapters as secondary literature sources. The papers identified in the search were subsequently analysed, and we included or excluded papers which met any of inclusion/exclusion criteria. All of the criteria are provided in Supplemental Material 1.

### ***Categorisation of the papers***

The selected papers provide information about currently used and developed quantitative methods and deal with the topic of quality issues in the use of central statistical monitoring. The selected papers were categorized in terms of their content into three categories: (1) ‘Central statistical monitoring processes’ focuses on central statistical monitoring in terms of quality issues and clinical trial workflow; (2) ‘Established central statistical monitoring methods’ refers to methods currently used within central statistical monitoring in different configurations; (3) ‘Innovative central statistical monitoring methods’ deals with techniques being developed for improving the central statistical monitoring methodology. In this context, innovation refers to techniques which extends beyond the basic methodology widely used in central statistical monitor’s daily practice.

### **Results**

The title and abstract review yielded initially 70 papers according to the applied query, with 24 included in this article as they met the pre-defined selection criteria. All of them were split into the pre-defined categories and it was possible for a paper to be classified into more than one category. Figure 1 depicts the whole literature selection process; and Table 2 shows included articles classified to defined categories.

<Figure 1>

<Table 2>

### *Central statistical monitoring processes*

Kirkwood et al. characterised central statistical monitoring as a more cost-effective alternative than full on-site monitoring, where data checks are performed by the coordinating centre, avoiding the need to visit sites as frequently.<sup>12</sup> Both approaches can co-exists in a proportion sufficient to improve the quality and reduce costs. In turn, Venet et al. described central statistical monitoring as a tool to detect abnormal patterns among clinical data and to help focus on sites which actually need to be monitored in detail.<sup>16</sup> Central statistical monitoring aims to avoid quality issues in the study conduct and the final outcome as well. In this context, quality can be defined as a set of product or process characteristics which arises from protecting patient well-being and providing a reliable assessment of investigated treatment.<sup>9</sup> Buyse et al. perceive central statistical monitoring as an improvement of quality management which relies on statistical indicators estimated on the basis of data collected across the whole trial,<sup>17</sup> in turn Pogue et al. consider central statistical monitoring more precisely as a part of quality assurance.<sup>18</sup>

Central statistical monitoring can be considered in different ways – from assuring data integrity to data analytics which helps to leverages information hidden in clinical data. This scope translates into the quality of the study outcome, which should be handled at all levels of the trial. As the expansion of clinical trials is progressing and requires more resources to be engaged, it is recommended to take measures respectively to the risk level. The paradigm of central statistical monitoring (and the entire risk based approach as well) needs to be

incorporated into the clinical trial workflow to ensure the synergy with current processes. The extension of this thought is available in Supplemental Material 2.

### ***Established central statistical monitoring methods***

The earliest central statistical monitoring methods typically relied on simple univariate statistical methods. These were used either individually or collectively to provide a more comprehensive understanding of the study conduct. Oba among others presented univariate approach within unsupervised learning to detect atypical data patterns.<sup>19</sup> He compared variable distribution between the sites and summarised them by descriptive statistics such as mean, median, standard deviation, skewness, or kurtosis. The distribution analysis was completed by data visualisation which is complementary to numeric results and facilitates their interpretation. Knepper et al. adopted a similar approach, investigating the study in terms of non-random error detection supported with data visualisation to make received statistics more understandable.<sup>20</sup>

Knepper et al. also conducted correlation check to examine relationships between variables and recognise unusual data patterns.<sup>20</sup> The method compares variables pairwise within a certain site. Kirkwood et al. compared the correlation structure between sites aiming to identify outlying ones.<sup>12</sup> In turn, Pogue et al. investigated the correlation of repeated measurements expressed by an intraclass correlation coefficient. This approach helps to find out fabricated data by flagging cases with high similarities of continuous measurements recorded multiple times.<sup>18</sup>

A battery of standard statistical tests is usually applied for comparing patients between sites.

<sup>21</sup> For example, Afroz et al. performed one-way ANOVA to compare two or more samples aided with post hoc analysis that helps to distinguish groups different from others. Another test they used was the chi-square goodness-of-fit test which evaluates two or more samples in



terms of similarity as well.<sup>22</sup> Kirkwood et al. recommend variance check as it helps to monitor sites in terms of repeated measurements.<sup>12</sup> Lower variance may indicate fabricated data at the site or patient level. Statistical tests are recommended for categorical data as well. Chi-square test of independence is applied in order to investigate patients' responses distribution among pre-defined categories. Kirkwood et al. suggest to use the test in the case of evaluation of safety, treatment compliance, and the primary efficacy endpoints.<sup>12</sup> In turn, Buyse et al. suggest to apply many tests rather than only a few of them and call it the essence of central statistical monitoring.<sup>17</sup> This approach is more efficient than single tests because it discloses knowledge hidden behind evidence accumulated across the entire study. Overall statistical testing makes better use of potential included in clinical data than focusing only on single metrics.

Descriptive statistics (frequency analysis) and statistical testing are utilized for digit preference analysis. According to Knepper et al.,<sup>20</sup> digit preference can be investigated by comparing digit frequency distribution of a single subject to other ones using chi-squared or Fisher exact test. Alternatively, the comparison can be performed by using mean and standard deviation in relation to a confidence interval. Another approach is based on the fact that every numeric result is considered as a sequence of digits. In the case of numbers with leading digits from 1 to 9, frequency of others should follow Benford's distribution. If clinical data are inconsistent with the distribution, it implies a fraudulent activity.<sup>12,22</sup> Oba proposed another approach to examine digit preferences in a particular measurement, i.e. volcano plot.<sup>19</sup> It is a scatter plot where  $y$ -axis represents a negative common logarithm of a raw  $p$ -value from testing digit preference, and the  $x$ -axis is a maximum difference in the frequency distribution of all digits between a single site versus all other sites. Then, the most outlying upper points on the plot indicate suspected sites (Supplemental Material 3, Figure S3-1).

Multiple statistical testing has another meaning for testing data consistency. Trotta et al. noticed that the abundance of statistical tests give too many results to analyse them one by one, therefore data aggregation is needed to facilitate the analysis.<sup>23</sup> All performed tests can be transformed into one metric known as a data inconsistency score which is a weighted geometric mean of the  $p$ -values of all tests comparing site  $i$  with all other sites, and can be calculated by the following formula:

$$DIS_i = \exp \left( \frac{1}{\sum_{j=1}^N w_j} \sum_{j=1}^N w_j \log p_{ij} \right) \quad (1)$$

where  $DIS_i$  is the data inconsistency score for site  $i$ ,  $w_j$  is a weight proportional to the correlation between the tests, and  $p_{ij}$  is a  $p$ -value of  $j^{\text{th}}$  test in site  $i$ . This operation transforms the outcome of all performed statistical tests into a one metric which makes sites easily comparable to each other, and helps to identify outliers considering the value obtained.<sup>17</sup> Actually, calculating the data inconsistency score is more complicated, but the simplified definition is sufficient for these considerations.<sup>23</sup>

The topic of dimensionality reduction is substantial for clinical data analysis as clinical trials handle many variables within a single study. These kinds of methods analyse leverage the information load form more than one variable. It is more useful in the context of inliers and outliers detection than univariate approach, ensuring more comprehensive analysis.<sup>19</sup> Pogue et al. consider multivariate methods in a greater extent,<sup>18</sup> i.e., from statistical summaries through statistical tests to purely graphical methods. Multivariate analysis closely relies on distance metrics which determine proximity (similarity) of data points. A particularly important metric within this category is the usage of Mahalanobis distance which Afroz et al. found as a multidimensional risk assessment method.<sup>22</sup> Mahalanobis distance works similarly to Euclidean distance, but raw data are normalized by the variance, which takes into account correlations between variables. Olsen et al. recognized Mahalanobis distance as the best

suited method for multivariate pattern and anomaly detection.<sup>24</sup> Venet et al. utilized principal component analysis to reduce dimensionality and presented results in the two-dimensional space.<sup>16</sup> Two dimensions are desired in order to facilitate visualisation of results in terms of outlier detection and make results more understandable.

The most comprehensible and attractive form of presenting study outcome is data visualisation. It is an ubiquitous method within cited aforementioned papers. Data visualisation usually completes performed analyses, which makes statistical concepts and raw data more tangible for a viewer via a graphical representation of the raw or aggregated data. Processes can be visualised as well, which helps to monitor them through the time. Van den Bor et al. presented the process of patient enrolment as a cumulative function,<sup>25</sup> and summarized with a boxplot in order to detect suspected sites. Afroz et al. compared sites by using boxplots considering Mahalanobis distance between them.<sup>22</sup> Buyse et al. utilized boxplots to compare  $p$ -values of the Cochran-Mantel-Haenszel test to the proportion of patient with errors.<sup>26</sup> Edwards et al. monitored a trial by using density plots which showed a distribution of the coefficient of variation for continuous variables (Supplemental Material 3, Figure S3-2).<sup>27</sup>

Cited researchers have already evaluated the effectiveness of various methods based on the type of clinical data they utilized. The last two columns in Table 3 show differences between implemented datasets. Some of them include cases labelled as fraud, which enables to evaluate the performance of tested techniques by using metrics like sensitivity, specificity, or detection rate.<sup>12,18,20,21,23,25</sup> However, Buyse et al. utilised labelled data differently. The initial data set was enriched with simulated erroneous data to examine the impact of their kind (random or systematic) on the study outcome.<sup>26</sup> On the other hand, analyses performed on unlabelled data just focused on flagging suspicious data points needed to be verified

individually.<sup>16,17,19,22,24,27</sup> This approach needed the expert knowledge or experience of the study team to build the medical context which helps to identify the potential cause.

### <Table 3>

#### *Innovative central statistical monitoring methods*

While the established central statistical monitoring methods perform satisfactorily, there is room for further improvement. The elusive nature of data errors and fraud, due to their subtlety, presents a challenge. Therefore, there is a need for expanding the range of tools further to make error and fraud detection more robust.

Churová et al. conducted anomaly detection by using clustering with seven distance metrics and their configurations.<sup>28</sup> The applied metrics are as follows: Canberra, Chebyshev, Cosine, Euclidean, Manhattan, Mahalanobis, and Minkowski. A distance matrix was determined per single metric and their derivatives as well. Calculated distances were examined by using univariate testing. Clustering was performed by using thresholds calculated individually for each metric twice: with predefined percentiles and following the interquartile range rule. As a result, a combination of Canberra, Manhattan and Mahalanobis gave the best performance, whereas Cosine and Chebyshev underperformed and were excluded from further analysis. Ultimately, Churova et al. achieved the optimal number of patients to investigate in the range 5.0-22.5% of total number of patients.

Zink et al. addressed development of the multivariate analysis methodology for inliers and outliers identification, and how single variables contribute to each multivariate outlier.<sup>29</sup> They proposed squared Mahalanobis distance which is proportional to Hotelling's  $T$ -squared distribution. The distribution is dedicated to handling multivariate hypothesis testing. On the basis of the distribution, a confidence interval was constructed as a criterion for outlier

identification. This approach can be supported with a scatterplot as a control chart for monitoring suspected values. The authors summarised findings by visualising the proportion of the squared Mahalanobis distance on a pareto plot (Supplemental Material 3, Figure S3-3). This helps to investigate outliers in terms of the most contributing variables, and hence points out areas for risk mitigation.

Hatayama and Yasui suggests Bayesian finite mixture model in order to detect atypical patterns across small and moderate studies.<sup>30</sup> The methods aims to estimate a mixture distribution which takes into account sub-populations in an overall dataset – i.e. fraudulent and non-fraudulent cases. Each of these categories has its own distribution, whereby the fraudulent one is a minority. Bayesian interference is used to estimate parameters of the models, but only the majority distribution is utilized by comparing averaged site results with it.

Desmet et al. proposed an application of the beta-binomial distribution for modelling event counts.<sup>31</sup> They utilized it for estimating a hybrid model which involves fraudulent and non-fraudulent sub-populations. The detection procedure relies on the following steps: (1) fit a beta-binomial model to the overall data, and (2) assign a  $p$ -value to each site and classify respectively to a significance level ( $\alpha_{cr}$ ). This criterion indicates which cases are fraudulent or not.

Desmet et al. proposed as well an application of the linear mixed-effects model to evaluate differences between single sites and the entire study.<sup>32</sup> They modelled a heterogeneous population composed of ‘normal’ and ‘abnormal’ observations. Each of those groups follows separate distributions with shifted means. The authors assumed a simple detection algorithm as a baseline for the model they investigated. According to the assumption, firstly, a hybrid model is estimated (including all of the data). A  $p$ -value is assigned to each observation. Then, observations are flagged respectively to assumed  $\alpha_{cr}$ . However, Desmet et al. noted that

the model is not always adequate for hybrid data and underperforms in terms of discrepancy detection. Therefore, they applied linear mixture-effects models to detect atypical sites across a study, which are dedicated to clustered data.  $P$ -values are assigned to each observation respective to the normal distribution estimated on the basis of the average variance calculated between a single site and the rest of them. Then, sites are flagged in relation to  $\alpha_{cr}$ .

Pogue et al. proposed logistic regression for fraud prediction by establishing cut-offs for the risk-scoring process.<sup>18</sup> The authors selected predictors and eliminated redundancy within them by using principal component analysis. They compared five three-variable models that have same two predictors. The predictors are associated with a risk level as a result of a statistical test performed for chosen attributes of the study. The risk was considered on a five-level scale in the range of 0-4. On this basis, the logistic model was estimated to predict the total risk level as a sum of scores assigned to predictors. Each model has its own cut-offs since it differs from the other with the third predictor. All models can be analysed at different risk levels and summarized in terms of a number of fraudulent cases, but it also indicates which risk level the dubious cases start from.

Zink et al. proposed funnel plots as a traffic-light system which indicates risk in terms of the duration of patient follow-up.<sup>33</sup> The authors analysed discontinuation rate versus total number of patients. This relationship is presented on a scatterplot with constant thresholds that indicate severity of the risk. This graph works similarly as a control chart to monitor a process (Supplemental Material 3, Figure S3-4). The authors moved away from an assumption that the risk is independent of the number of patients to apply asymptotic thresholds. The higher risk is observed for smaller sites and it gradually decreases proportionate to increasing number of patients. Zink et al. proposed how to calculate the thresholds based on the asymptotic confidence interval of the event rate. This method was presented through the prism of two perspectives: binary outcomes, where events are counted zero-one independently to their

repetition, and count-type outcomes, where the number of event repetitions is taken into account. In the first case, the confidence interval is determined on the basis of the observed event rate at the site, and the second one utilizes the Poisson distribution. Funnel plots turned out to be a useful graphical technique that considers changes in a sample size, patients exposure, and risk monitored through the time.

<Table 4>

## **Discussion**

The objectives of the scoping review were to define an established framework for the central statistical monitoring process in the context of the study conduct, and identify quantitative methods underlying this approach including innovations. The invention of the method of central statistical monitoring is a response to regulators requirements and guidelines which recommend the risk-based approach for monitoring of clinical trials. Central statistical monitoring is in the process of continuous improvement at conceptual, strategic and operational levels. It results in development of monitoring processes aided by data mining algorithms.

Researchers and practitioners perceive central statistical monitoring as a part of the quality management in clinical trials. This point of view is consistent with regulators' recommendations as they require the implementation of a quality management system across the entire clinical trial workflow. The pivotal factor that contributes to the quality of the final study outcome is data consistency. Consistency of the data is considered equivalent to the data quality and can be measured by using the wide range of quantitative methods aided with data visualisation.

The outcome of well-performed central statistical monitoring analyses is a roadmap for high-quality decision-making. The reliability of the results determine the rationality of decisions and the future of the clinical trials in the long terms. The data-driven approach is not a substitute of decision-making based on human perception but a support in making optimal decisions. Data-driven decision-making is consistent with the idea of risk-based approach which involves the risk-based quality management strategy with risk-based monitoring as its key component. This strategy aims to manage tasks proportional to the perceived risk level, which allows the organisation to avoid unnecessary resource exploitation. Then, it makes the study conduct more efficient, which creates value for stakeholders of the drug development process and rises its quality. As a result, the risk-based approach forms a new dimension of the quality management.

To implement a risk-based quality management strategy effectively, it is necessary to incorporate central statistical monitoring into the clinical trial workflow. Its place in the process results from the existence of issues occurring in the data generation processes that fail to meet those defined in the study protocol. Discrepancies identified in clinical data provide information which is a starting point for determining corrective actions. They can be caused by fraud or just simple mistakes committed during the course of the study, but its signals are hidden among the collected data. These cases tend to leave some clues which may be identified by using data mining algorithms.

Central statistical monitoring can be carried out with a battery of established methods, which are well-known to practitioners. These methods were grouped into categories such as statistical testing, descriptive statistics, data inconsistency score evaluation, correlation analysis, digit preferences check, multivariate analysis, and data visualisation. These methods are applied in various configurations, as their outcomes complement each other. Their usage is recognised as a must-have within the framework of central statistical review. In our opinion



the biggest potential lies in multivariate techniques as clinical trials are analysed in the multidimensional space outlined by variables involved in the study. Associations between these variables are not obvious for a viewer at first glance. Therefore, multivariate methods seem to be more powerful<sup>19</sup> compared to the univariate approach which might be insufficient<sup>24</sup>. Multivariate analysis discloses hidden data patterns and trends resulting from the natural multicollinearity in clinical data<sup>18,29</sup>. Data visualisation is a powerful field as well, being ubiquitous among cited articles, thus proving its utility. Graphical representation of the data provides easy to interpret summaries of performed statistical analyses. Graphics can be shared with other people, without need for detailed statistical knowledge, in order to communicate analysis results in an accessible way.

Although some methods are already established, their performance is insufficient to extract all valuable information included in the collected data. The need for innovations in the central statistical monitoring methodology is highlighted in the literature.<sup>12,17</sup> A part of the methods was tested in simulated data which do not fully reflect the real nature of clinical data. It is recommended to improve the current methodology by refining available algorithms on data from actual studies. Cited researchers addressed the development of methods using multivariate analysis, mixed-effect models, data visualisation etc. Of course, it does not exhaust directions of further research but confirms the aforementioned statement about method usability.

Our objective in this scoping review was to consider the central statistical monitoring methods currently available in the literature. The publications mainly show the improvement in the data through finding these errors. Proving the value of central statistical modelling compared to other approaches, for example on-site monitoring, is outside the remit of this study but has been discussed in other sources<sup>8,12,16,34-36</sup>. Moreover, this approach is recommended by both good clinical practice and regulatory agencies.<sup>1,2,6</sup>

This scoping review summarises the concept of central statistical monitoring and identified methods used in this field or being developed. The extracted knowledge was consolidated into one consistent image of central statistical monitoring focusing on quality issues. The method identification showed what is the current state of the central statistical monitoring methodology development and what directions are noteworthy for further research on data mining algorithms.

## **Acknowledgements**

We would like to thank all members of the joint the PSI/EFSP/ASA-Biop Special Interest Group: Central Statistical Monitoring & Quality Tolerance Limits for sharing their experiences of CSM in drug development both from a pharmaceutical industry and academic perspective and for supporting the need to conduct this literature research. We would also like to thank GlaxoSmithKline plc (GSK) for supporting Maciej Fronc in his PhD research and the authors in the research used in this paper.

## **Funding**

Sharon Love was supported by the Medical Research Council grant numbers MC\_UU\_00004/08.

## **Declaration of Conflicting Interests**

The authors declare that there is no conflict of interest.

## **References**

1. Food and Drug Administration (FDA). *Guidance for Industry Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring Guidance for Industry*

- Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring*, <https://www.fda.gov/media/116754/download> (2013).
2. European Medicines Agency (EMA). *Reflection paper on risk based quality management in clinical trials*, [https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en.pdf) (2013).
  3. Agrafiotis DK, Lobanov VS, Farnum MA, et al. Risk-based Monitoring of Clinical Trials: An Integrative Approach. *Clin Ther* 2018; 40: 1204–1212.
  4. Baigent C, Harrell FE, Buyse M, et al. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clin Trials* 2008; 5: 49–55.
  5. George SL, Buyse M. Data fraud in clinical trials. *Clin Investig (Lond)* 2015; 5: 161–173.
  6. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2)*, [https://database.ich.org/sites/default/files/E6\\_R2\\_Addendum.pdf](https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf) (2016).
  7. TransCelerate Biopharma Inc. RBM Interactive Guide, <https://www.transceleratebiopharmainc.com/rbminteractiveguide/what-is-risk-based-monitoring-rbm/introduction/> (2012, accessed 10 January 2023).
  8. Wilson B, Provencher T, Gough J, et al. Defining a Central Monitoring Capability: Sharing the Experience of TransCelerate BioPharma’s Approach, Part 1. *Ther Innov Regul Sci* 2014; 48: 529–535.
  9. Timmermans C, Venet D, Burzykowski T. Data-driven risk identification in phase III clinical trials using central statistical monitoring. *Int J Clin Oncol* 2016; 21: 38–45.
  10. Fronc M, Jakubczyk M. From business to clinical trials: a systematic review of the

literature on fraud detection methods to be used in central statistical monitoring.

*Przeegląd Stat* 2022; 69: 1–31.

11. Smith R. Time to assume that health research is fraudulent until proven otherwise?, <https://blogs.bmj.com/bmj/2021/07/05/time-to-assume-that-health-research-is-fraudulent-until-proved-otherwise/> (2022, accessed 16 January 2023).
12. Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. *Clin Trials* 2013; 10: 783–806.
13. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *Int J Soc Res Methodol Theory Pract* 2005; 8: 19–32.
14. PRISMA for Scoping Reviews, <http://www.prisma-statement.org/Extensions/ScopingReviews> (2018).
15. Munn Z, Peters MDJ, Stern C, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018; 18: 143.
16. Venet D, Doffagne E, Burzykowski T, et al. A statistical approach to central monitoring of data quality in clinical trials. *Clin Trials* 2012; 9: 705–713.
17. Buyse M, Trotta L, Saad ED, et al. Central statistical monitoring of investigator-led clinical trials in oncology. *Int J Clin Oncol* 2020; 25: 1207–1214.
18. Pogue JM, Devereaux PJ, Thorlund K, et al. Central statistical monitoring: Detecting fraud in clinical trials. *Clin Trials* 2013; 10: 225–235.
19. Oba K. Statistical challenges for central monitoring in clinical trials: a review. *Int J Clin Oncol* 2016; 21: 28–37.
20. Knepper D, Lindblad AS, Sharma G, et al. Statistical Monitoring in Clinical Trials: Best Practices for Detecting Data Anomalies Suggestive of Fabrication or Misconduct. *Ther Innov Regul Sci* 2016; 50: 144–154.

21. de Viron S, Trotta L, Schumacher H, et al. Detection of Fraud in a Clinical Trial Using Unsupervised Statistical Monitoring. *Ther Innov Regul Sci* 2022; 56: 130–136.
22. Afroz MA, Schwarber G, Bhuiyan MAN. Risk-based centralized data monitoring of clinical trials at the time of COVID-19 pandemic. *Contemp Clin Trials* 2021; 104: 106368.
23. Trotta L, Kabeya Y, Buyse M, et al. Detection of atypical data in multicenter clinical trials using unsupervised statistical monitoring. *Clin Trials* 2019; 16: 512–522.
24. Olsen MH, Hansen ML, Safi S, et al. Central data monitoring in the multicentre randomised SafeBoosC-III trial – a pragmatic approach. *BMC Med Res Methodol* 2021; 21: 1–10.
25. van den Bor RM, Vaessen PWJ, Oosterman BJ, et al. A computationally simple central monitoring procedure, effectively applied to empirical trial data with known fraud. *J Clin Epidemiol* 2017; 87: 59–69.
26. Buyse M, Squifflet P, Coart E, et al. The impact of data errors on the outcome of randomized clinical trials. *Clin Trials* 2017; 14: 499–506.
27. Edwards P, Shakur H, Barnetson L, et al. Central and statistical data monitoring in the Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage (CRASH-2) trial. *Clin Trials* 2014; 11: 336–343.
28. Churová V, Vyškovský R, Maršálová K, et al. Anomaly detection algorithm for real-world data and evidence in clinical research: Implementation, evaluation, and validation study. *JMIR Med Informatics* 2021; 9: 1–17.
29. Zink RC, Castro-Schilo L, Ding J. Understanding the influence of individual variables contributing to multivariate outliers in assessments of data quality. *Pharm Stat* 2018; 17: 846–853.
30. Hatayama T, Yasui S. Bayesian central statistical monitoring using finite mixture

- models in multicenter clinical trials. *Contemp Clin Trials Commun* 2020; 19: 100566.
31. Desmet L, Venet D, Doffagne E, et al. Use of the Beta-Binomial Model for Central Statistical Monitoring of Multicenter Clinical Trials. *Stat Biopharm Res* 2017; 9: 1–11.
  32. Desmet L, Venet D, Doffagne E, et al. Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. *Stat Med* 2014; 33: 5265–5279.
  33. Zink RC, Dmitrienko A, Dmitrienko A. Rethinking the Clinically Based Thresholds of TransCelerate BioPharma for Risk-Based Monitoring. *Ther Innov Regul Sci* 2018; 52: 560–571.
  34. Adams A, Adelfio A, Barnes B, et al. Risk-Based Monitoring in Clinical Trials: 2021 Update. *Ther Innov Regul Sci* 2023; 57: 529–537.
  35. de Viron S, Trotta L, Steijn W, et al. Does Central Monitoring Lead to Higher Quality? An Analysis of Key Risk Indicator Outcomes. *Ther Innov Regul Sci* 2023; 57: 295–303.
  36. de Viron S, Trotta L, Steijn W, et al. Does Central Statistical Monitoring Improve Data Quality? An Analysis of 1,111 Sites in 159 Clinical Trials. *Ther Innov Regul Sci* 2024; 58: 483–494.

Table 1. Keywords used in the query

<b>Part of the query</b>	<b>Purpose (research axis)</b>
(“clinical trials” AND (“central* monitoring” OR “statistical monitoring”))	Application field
“quality” OR “management”	Process and quality context
(“machine learning” OR “data mining” OR “artificial intelligence” OR statistic* OR econometric*)	Expected methodology
(fraud* OR manipulation OR fabrication OR misconduct OR “systematic error”)	Problem to be solved

Table 2. Total summary of included papers

No.	Author	Year	DOI	CSM processes	Established CSM methods	Innovative CSM methods
1.	Afroz et al.	2021	10.1016/j.cct.2021.106368			+
2.	Churova et al.	2021	10.2196/27172		+	
3.	Cragg et al.	2021	10.1177/1740774520976561	+		
4.	de Viron et al.	2021	10.1007/s43441-021-00341-5			+
5.	Li et al.	2021	10.1097/AOG.0000000000004513	+		
6.	Olsen et al.	2021	10.1186/s12874-021-01344-4			+
7.	Buyse et al.	2020	10.1007/s10147-020-01726-6	+		+
8.	Hatayama & Yasui	2020	10.1016/j.conctc.2020.100566	+	+	
9.	Trotta et al.	2019	10.1177/1740774519862564			+
10.	Zink, Castro-Schilo et al.	2018	10.1002/pst.1903		+	
11.	Dmitrienko et al.	2018	10.1177/2168479017738981		+	
12.	Buyse et al.	2017	10.1177/1740774517716158			+
13.	Desmet et al.	2017	10.1080/19466315.2016.1164751		+	
14.	van der Bor	2017	10.1016/j.jclinepi.2017.03.018			+
15.	Herson	2016	10.1007/s10147-015-0876-6	+		
16.	Knepper et al.	2016	10.1177/2168479016630576			+
17.	Oba	2016	10.1007/s10147-015-0914-4			+
18.	Timmermans et al.	2016	10.1007/s10120-015-0533-9	+		
19.	Timmermans et al.	2016	10.1007/s10147-015-0877-5	+		
20.	Desmet et al.	2014	10.1002/sim.6294		+	
21.	Edwards et al.	2014	10.1177/1740774513514145			+
22.	Kirkwood et al.	2013	10.1177/1740774513494504	+		+
23.	Pogue et al.	2013	10.1177/1740774512469312		+	+
24.	Venet et al.	2012	10.1177/1740774512447898	+		+



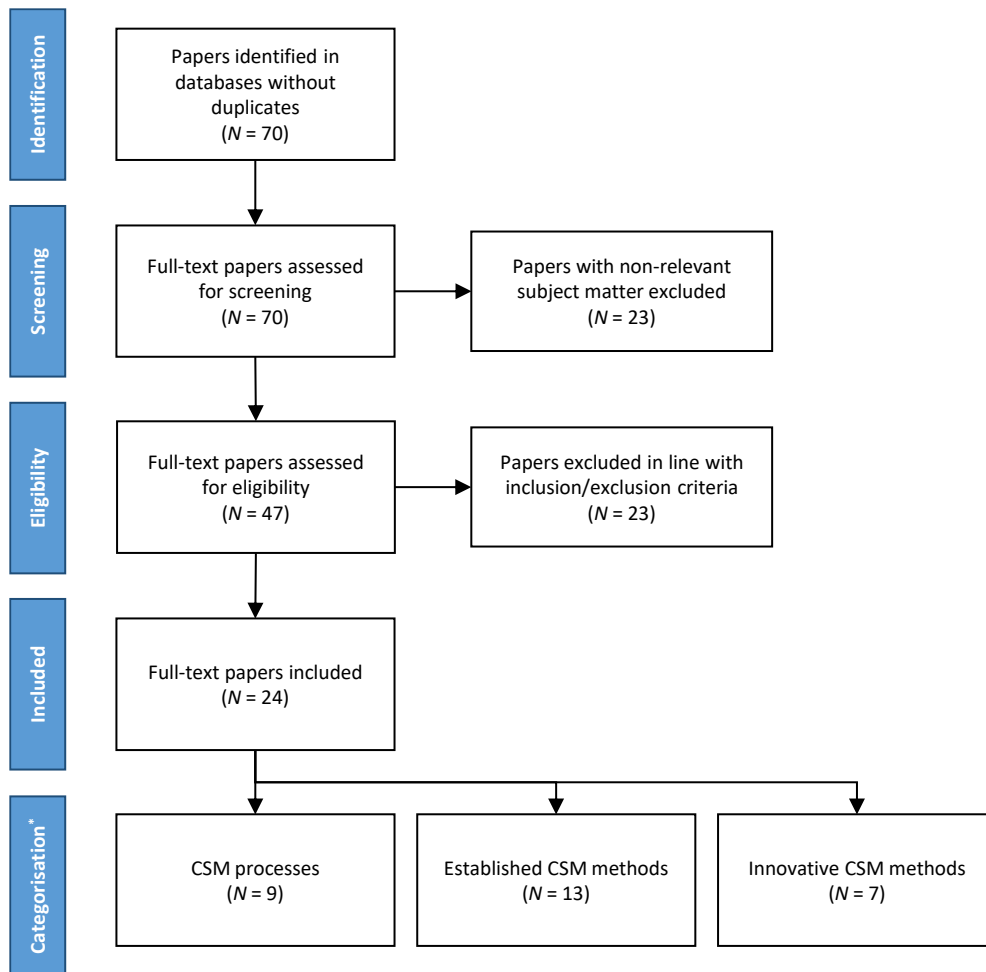
Table 3. Summary of papers in terms of basic method usage

No.	Author	Year	Statistical testing	Descriptive statistics	DIS	Correlation analysis	Digit preference	Multivariate analysis	Data visualization	Clinical data	Fraudulent cases
1.	de Viron et al.	2022	+		+				+	Original	Known
2.	Afroz et al.	2021	+	+			+	+	+	Simulated	Unknown
3.	Olsen et al.	2021		+				+	+	Original	Unknown
4.	Buyse et al.	2020	+		+				+	Original	Unknown
5.	Trotta et al.	2019	+		+				+	Modified	Known
6.	Buyse et al.	2017	+						+	Modified	Known
7.	van den Bor et al.	2017				+	+		+	Original	Known
8.	Knepper et al.	2016		+		+	+	+	+	Modified	Known
9.	Oba	2016		+			+	+	+	NA	Unknown
10.	Edwards et al.	2014		+					+	Original	Unknown
11.	Kirkwood et al.	2013	+	+		+	+		+	Modified	Known
12.	Pogue et al.	2013	+	+		+	+	+	+	Original	Known
13.	Venet et al.	2012						+	+	Original	Unknown

Note: ‘+’ means the method is present in the paper; and ‘NA’ means the information is not available. Three categories of clinical data were distinguished: ‘original’ – from real clinical trials without any modifications; ‘modified’ – original clinical data partially modified; and ‘simulated’ – fully modelled on original clinical data. And two categories of fraudulent cases were defined: ‘known’ – fraudulent cases are pointed out in the dataset; and ‘unknown’ – data unlabelled in terms of fraudulent activity.

Table 4. Summary of papers in terms of methods in development

No.	Authors	Year	Method
1.	Churová et al.	2021	Clustering with 7 different distance metrics in order to identify a combination of them with the best anomaly detection
2.	Hatayama & Yasui	2020	Bayesian finite mixture model for atypical sites detection
3.	Zink, Castro-Schilo, et al.	2018	Squared Mahalanobis distances for outlier identification and assessment of the individual variable's contribution to multivariate outliers
4.	Zink, Dmitrienko, et al.	2018	Funnel plots for monitoring discontinuation rate
5.	Desmet et al.	2017	Statistical modelling by using beta-binomial distribution for detection of atypical sites in terms of probability of some events
6.	Desmet et al.	2014	Linear mixed-effects models for differences identification between sites
7.	Pogue et al.	2013	Logistic regression for risk prediction



\* The total number of papers from this category is greater than the total number of included articles because a part of them was classified into more than one category.

Figure 1. Stages of search process for the review

# Supplemental Material 1

---

## Central statistical monitoring in clinical trial management: a scoping review

Maciej Fronc, Michał Jakubczyk, Sharon B. Love, Susan Talbot, Timothy Rolfe

### Details on inclusion and exclusion criteria

The papers identified in the search were subsequently analysed, and we included papers which met any of the following inclusion criteria:

- Papers addressing remote monitoring strategies of CTs that outline the theoretical framework for CSM;
- Papers presenting the currently used CSM methodology;
- Research on development of the CSM methodology;
- Papers decomposing the concept of the risk-based approach to clinical trials;
- Papers addressing the issue of misconduct in CTs in the context of quality management.

At least one reason sufficed for the article to be excluded:

- Papers with repeated subject matter. The one which contributed more to the development or explanation of CSM was selected;
- Research on on-site monitoring excluding centralized approach;
- Topics from the list of inclusion criteria are discussed only superficially;
- Papers not detailed in making data-driven decisions.
- All choices were discussed between authors at every stage of the reviewing process, and the final decision on selection of the articles was made by consensus.

## Supplemental Material 2

### Central statistical monitoring in clinical trial management: a scoping review

Maciej Fronc, Michał Jakubczyk, Sharon B. Love, Susan Talbot, Timothy Rolfe

#### CSM processes (contd.)

Timmermans et al. distinguished four areas of CTs which contribute to the quality level of the final outcome,<sup>1</sup> i.e.: study design, study conduct, analysis of the results and reporting. ICH recommends sponsors to implement a quality management system across all stages of the trial process.<sup>2</sup> The system should involve measures that have an impact on quality of CTs from their early stages (Figure S2-1).

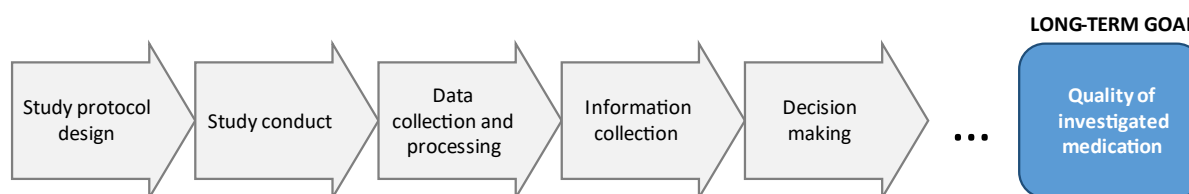


Figure S2-1. Stages of CTs impacted by quality management processes

The stage of data collection and processing is essential as it interacts indirectly with quality of the final product. Data quality is strictly dependent on data consistency which can be measured across all study sites by performing statistical analyses.<sup>3</sup> Li et al. addressed the issue of data integrity – i.e. overall accuracy, completeness, and consistency of data – based on results of clinical trials published in scientific journals.<sup>4</sup> They focused on statistical assessment of misconduct in CTs as a source of data integrity issues. The researchers proposed two approaches of data integrity assessment (DIA) depending on clinical data availability: DIA with individual data where all results are available, and DIA without individual data where detailed results of participants are unknown and only aggregated values are shared. The first approach utilizes methods that compare a single patient with a general tendency across the study. Even calculations are not necessary, and the whole assessment might be limited to data visualization and simple comparison of single values. In turn, the second approach based on statistical testing which provides general information about sites and determines differences between them by verifying statistical hypotheses.

The prospect of risk applied to CTs suggest to draw attention to ‘things that really matter’ within study conduct.<sup>3</sup> Regulators recommend conducting risk assessment, which is

particularly important, as it determines the monitoring strategy as a part of quality management. Then single subjects or sites are investigated in detail proportionate to risk distributed across the whole study. The application of RBQM aims to avoid overcharging the monitoring strategy and improve its efficiency.<sup>5</sup> Monitoring conducted respectively to study risk is called risk-based monitoring (RBM) and it is a component of the RBQM concept. RBM capitalises on monitoring conducted in the centralized way based on remote evaluation of study conduct and data capture.<sup>1</sup> This approach supports operationalization of the RBM concept.<sup>3</sup> What is more, Timmermans et al. recognized RBM as a new paradigm of quality control.<sup>6</sup> Effects of RBM guides researchers in applying reasonable actions seeking to deliver an expected quality level of a final product. Moreover, Hatayama and Yasui appeal to Pareto principle (otherwise known as the 80:20 rule)<sup>7</sup> as pivotal concept of quality management,<sup>8</sup> which is consistent with the CSM concept. While intuition may suggest that irregularities are limited to a minority of study sites, this minority often contributes to the majority of errors in study conduct. This point of view connects CSM with quality management within an indirect relation (Figure S2-2).

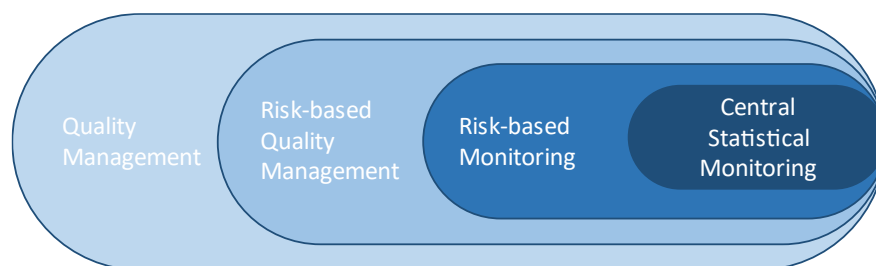


Figure S2-2. Linkage of the concepts underlying the centralised monitoring

*Where CSM is located in the CT workflow and how it operates?*

The goal of a clinical trial protocol is to provide a suitably homogeneous population in which to reliably characterise the effects of the intervention under investigation. As such, there should not be any major data discrepancies as all sites follow the same procedures within the single study. Any discrepancies could be minor but ultimately may increase the risk of study failure. Buyse et al. indicated typical causes of data issues such as fraud, data tampering, sloppiness, or miscalibration.<sup>3</sup> All of them might jeopardise the study outcome. Although it is difficult to capture them with naked eye, these causes leave signals detectable by applied algorithms. The signals are usually related to missing values or wrong number of records (reporting), atypical distribution of a feature (data tendency), observable change of a parameter over time (visit-to-visit evolution), or atypically reported visit days (visit date).<sup>6</sup> Signals and their causes are closely linked making them interdependent (Figure S2-3).

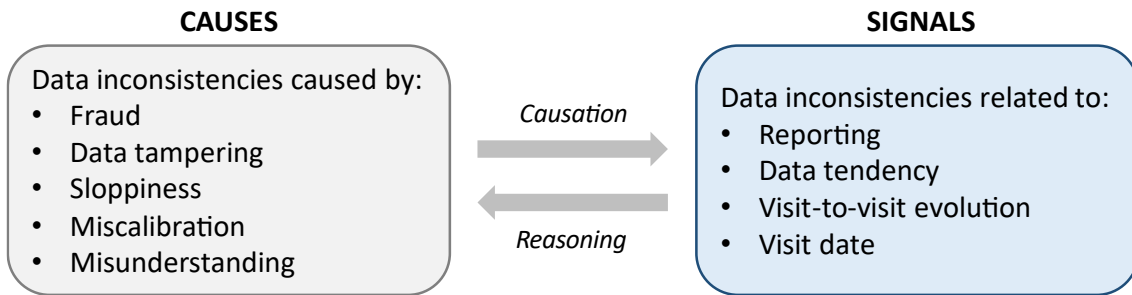


Figure S2-3. Interdependence of causes and signals

Sloppiness and miscalibration happen unintentionally, however, fraud and data tampering are committed consciously. Cragg et al. noticed that two thirds of articles about RBM they reviewed concerned CSM and a half of them addressed detection of fabricated data in CTs.<sup>5</sup> It shows a popularity of this methodology resulting from the scale of fraudulent activity.

Fraud easily evades human cognition but is detectable by using statistical approach.<sup>9</sup> Humans are unable to come up with credible data due to their incomplete knowledge of how real data are generated.<sup>4</sup> Herson defined two strategies for dealing with fraud – defensive and offensive ones.<sup>10</sup> Defensive strategy works in the spirit of quality by design philosophy, and it aims to prevent fraud by proper study planning. In turn, offensive strategy involves reacting to fraud detected through CSM, and implementing mitigation of the risk (Figure S2-4).

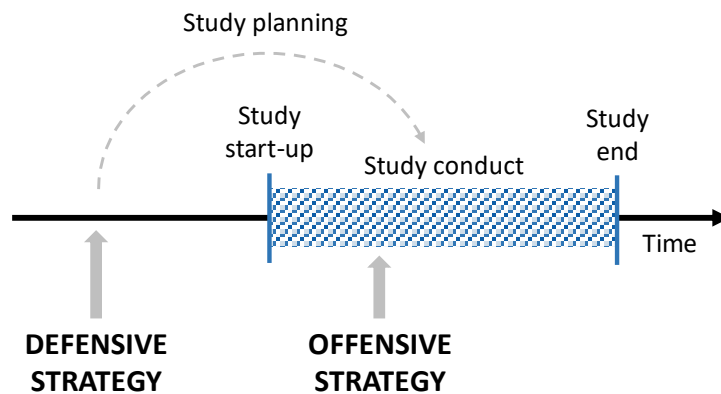


Figure S2-4. Timeline of defensive and offensive strategies application

Application of CSM decreases risk across the study. Identified KRIs require further investigation and application of preventive measures. KRIs are specific to the study, therefore they need to be customized to conduct targeted monitoring. Timmermans et al. even recommends not to have any prior assumption before statistical analysis to prevent omitting any risks.<sup>6</sup> Signals detected by using statistical tools are a premise for targeted source data

review (SDR) and if warranted, source data verification (SDV) at sites marked with ‘red flags’. CSM facilitates decision-making on the necessity to conduct SDV/SDR.<sup>5</sup> The outcome of CSM helps to avoid a need for full SDV/SDR at all sites across the study (Figure S2-5).

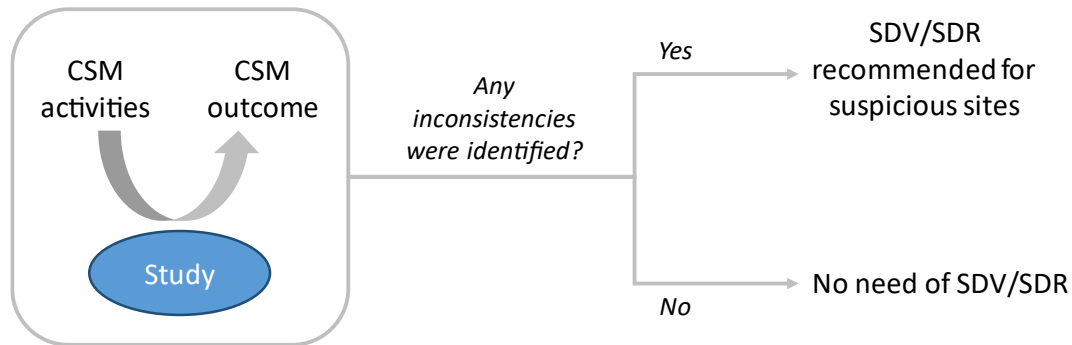


Figure S2-5. Decision-making involving the CSM outcome

CSM is a likely cheaper alternative to routine on-site monitoring of all sites within a study. Apart from the advantages of CSM, there are also some limitations. The currently used methods are selective and underperform in the case of small trials.<sup>5,11</sup> What is more, developed algorithms are often tested on fabricated data instead of using real-world data. Data are fabricated for this purpose with certain assumptions which might turn out to be incorrect and not reflect the inner nature of clinical data. Then, the algorithm may underperform and give unsatisfactory results. Kirkwood et al. encourage other researchers to refine these algorithms on their own datasets, and especially from actual CTs.<sup>11</sup>

## References

1. Timmermans C, Venet D, Burzykowski T. Data-driven risk identification in phase III clinical trials using central statistical monitoring. *Int J Clin Oncol* 2016; 21: 38–45.
2. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Guideline Good Clinical Practice E6(R2)*, [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf) (2016).
3. Buyse M, Trotta L, Saad ED, et al. Central statistical monitoring of investigator-led clinical trials in oncology. *Int J Clin Oncol* 2020; 25: 1207–1214.
4. Li W, Bordewijk EM, Mol BW. Assessing research misconduct in randomized controlled trials. *Obstet Gynecol* 2021; 138: 338–347.



5. Cragg WJ, Hurley C, Yorke-Edwards V, et al. Dynamic methods for ongoing assessment of site-level risk in risk-based monitoring of clinical trials: A scoping review. *Clin Trials* 2021; 18: 245–259.
6. Timmermans C, Doffagne E, Venet D, et al. Statistical monitoring of data quality and consistency in the Stomach Cancer Adjuvant Multi-institutional Trial Group Trial. *Gastric Cancer* 2016; 19: 24–30.
7. Hatayama T, Yasui S. Bayesian central statistical monitoring using finite mixture models in multicenter clinical trials. *Contemp Clin Trials Commun* 2020; 19: 100566.
8. Juran J, Godfrey A. *Juran's quality handbook*. wyd. 5. New York: McGraw-Hill Professional Publishing, 1998.
9. Fronc M, Jakubczyk M. From business to clinical trials: a systematic review of the literature on fraud detection methods to be used in central statistical monitoring. *Przełqd Stat* 2022; 69: 1–31.
10. Herson J. Strategies for dealing with fraud in clinical trials. *Int J Clin Oncol* 2016; 21: 22–27.
11. Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. *Clin Trials* 2013; 10: 783–806.

# Supplemental Material 3

## Central statistical monitoring in clinical trial management: a scoping review

Maciej Fronc, Michał Jakubczyk, Sharon B. Love, Susan Talbot, Timothy Rolfe

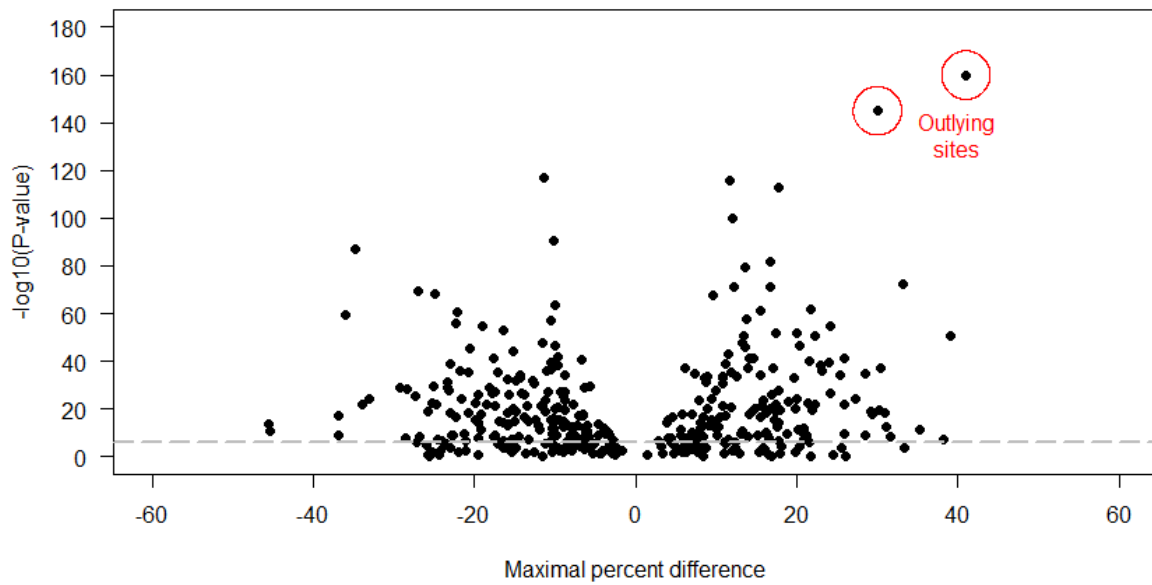


Figure S3-1. A volcano plot showing the distribution of terminal digit preference across all sites

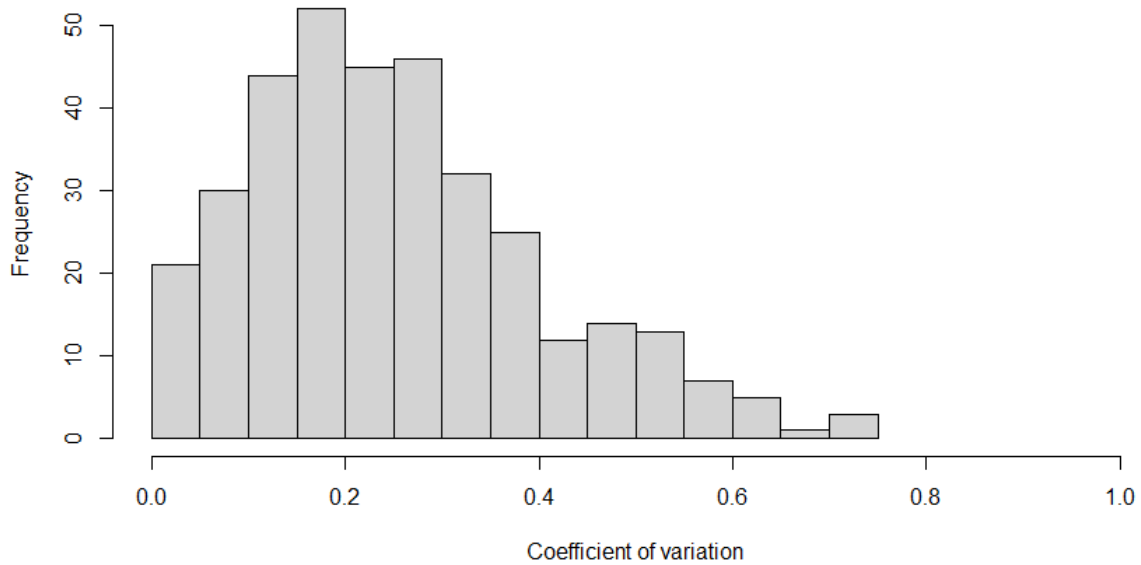
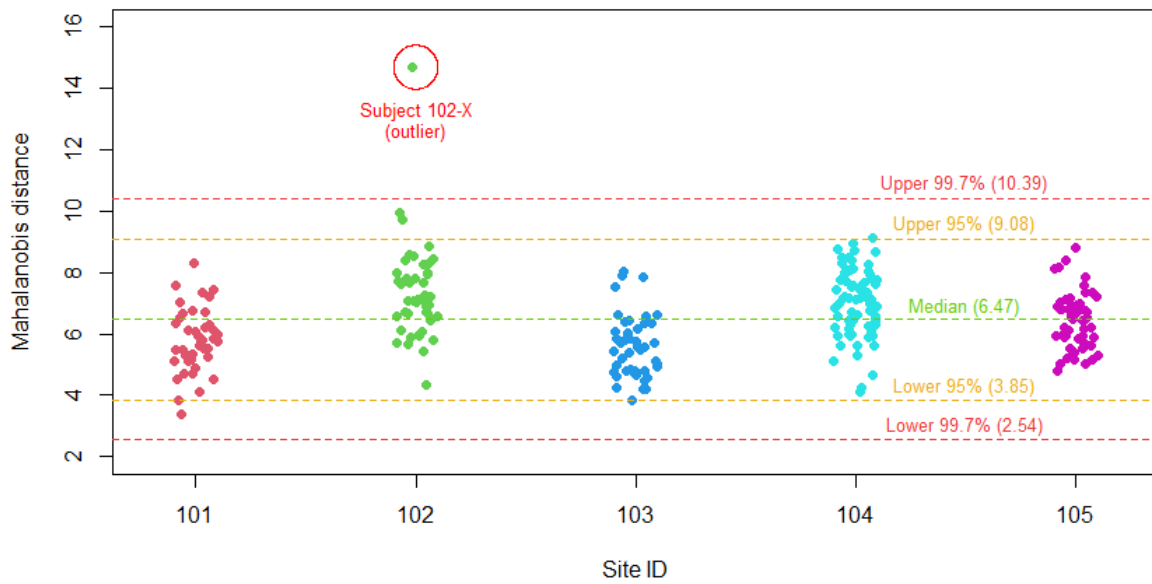


Figure S3-2. A distribution of the coefficient of variation for continuous variable

a)



b)

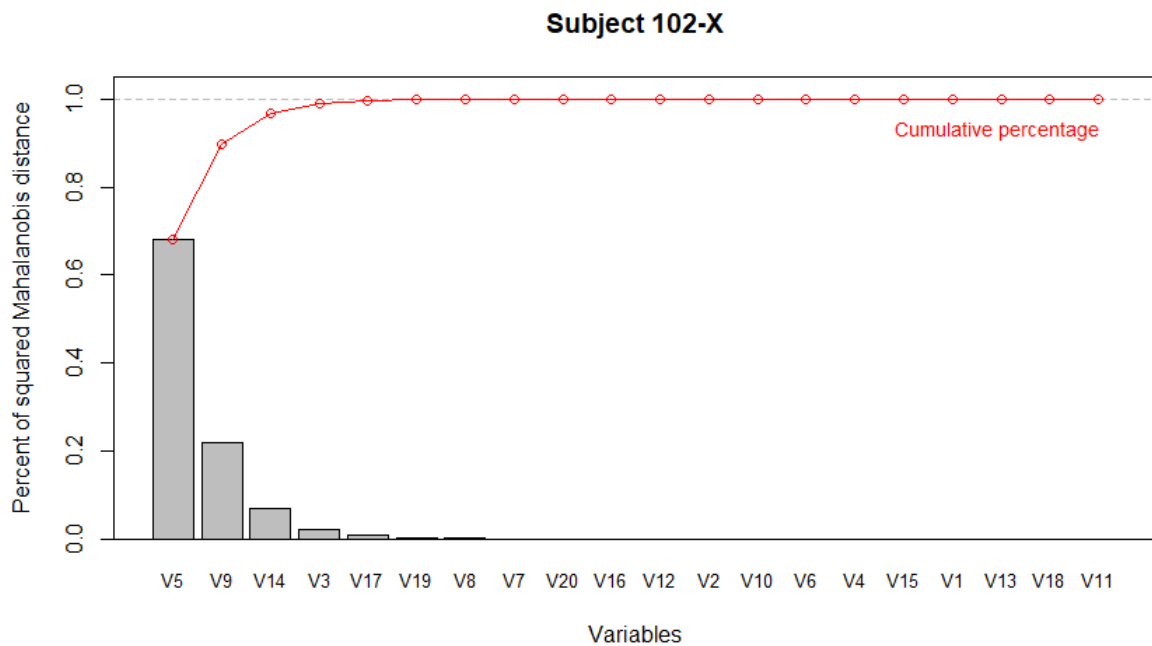


Figure S3-3. Outlier detection in terms of the Mahalanobis distance: (a) A scatterplot of Mahalanobis distances by site. Control limits (red and yellow dashed lines) are based on 95% and 99.7% Upper limits help to identify outliers, while lower limits help to identify inliers. Only one outlier (subject 102-X) was flagged. (b) A Pareto plot for patient 102-X showing the contribution of covariates to the outlier.

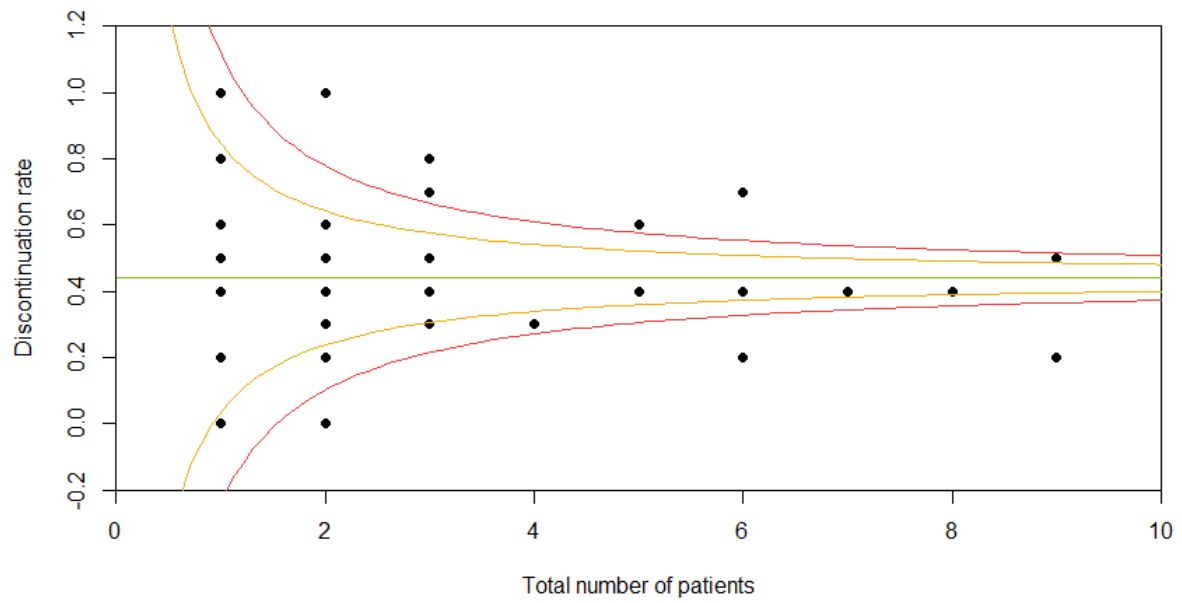


Figure S3-4. A funnel plot for the rate of discontinuation. The green line is the average discontinuation rate, while yellow and red lines are control limits based on 95% and 99.7% confidence interval.