Using central monitoring to detect data fabrication in multi-center clinical trials

Rutger van den Bor (r.m.vandenbor@umcutrecht.nl)

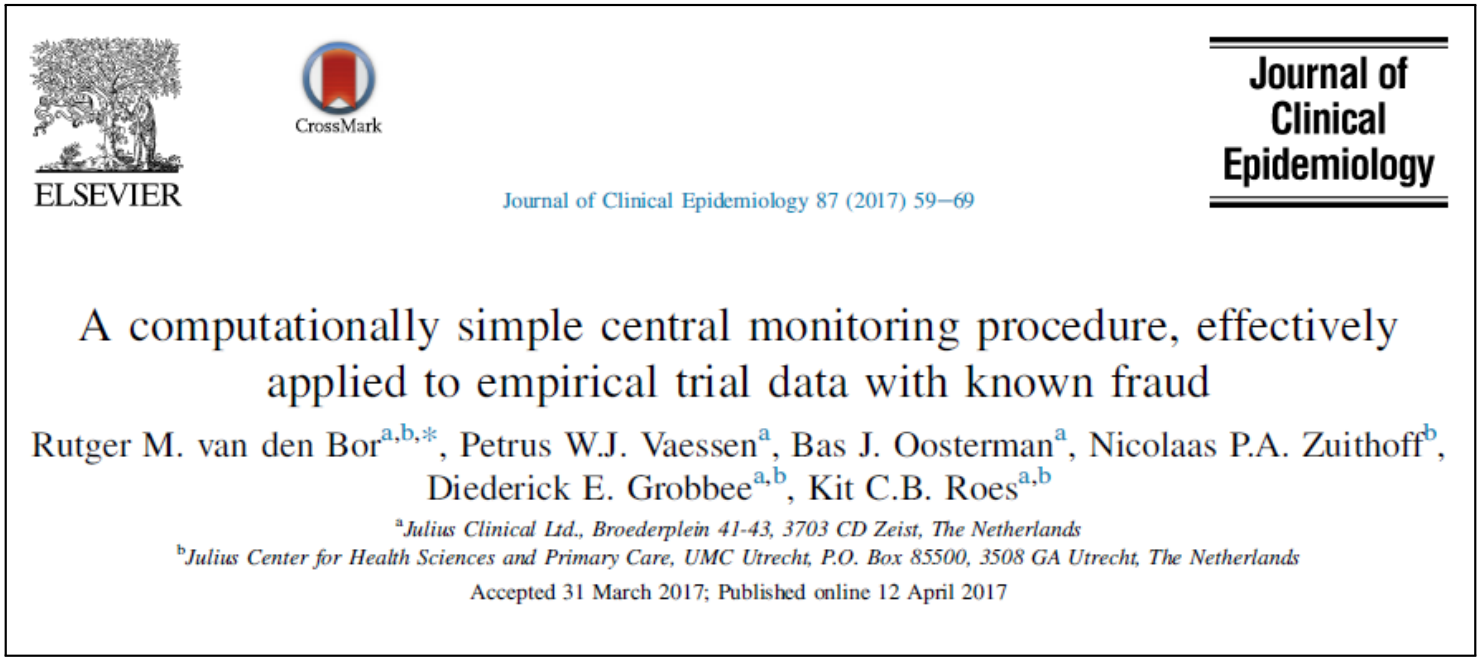
Department of Data Science and Biostatistics
Julius Center for Health Sciences and Primary Care, UMC Utrecht

PSDM event on risk-based approaches, 2021

Disclaimer

Views/opinions expressed in this presentation are solely those of the presenter and not necessarily those of UMC Utrecht or Julius Clinical

Publication



ELSEVIER

CrossMark

Journal of Clinical Epidemiology 87 (2017) 59–69

Journal of Clinical Epidemiology

A computationally simple central monitoring procedure, effectively applied to empirical trial data with known fraud

Rutger M. van den Bor[a,b,*], Petrus W.J. Vaessen[a], Bas J. Oosterman[a], Nicolaas P.A. Zuithoff[b], Diederick E. Grobbee[a,b], Kit C.B. Roes[a,b]

[a]*Julius Clinical Ltd., Broederplein 41-43, 3703 CD Zeist, The Netherlands*

[b]*Julius Center for Health Sciences and Primary Care, UMC Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands*

Accepted 31 March 2017; Published online 12 April 2017

Note: `R` scripts available in the online appendix

Fraud/data fabrication in clinical trials

Prevalence:

- Difficult to estimate
- Generally assumed to be low

Impact on trial results:

- Depends on trial characteristics and nature/extent of fraud
- Impact of one fraudulent center in a large multi-center double blinded RCT?

Fraud/data fabrication in clinical trials

However:

> *"Even isolated and small amounts of fraud within a trial can cause significant doubts about its conclusions and have the potential to lead to a lack of public confidence for the clinical trial process in general."*
>
> Pogue et al. (2013, p.226)

> *"We condemn all data fabrication. [...] Confidence in the integrity of the trial and its results is essential to every trial. If, through intentional or inadvertent actions, that confidence is impaired, not only have the participants and potentially others in the community been harmed, the trial loses its rationale, which is to influence science and medical practice."*
>
> Friedman et al. (2015, p.42)

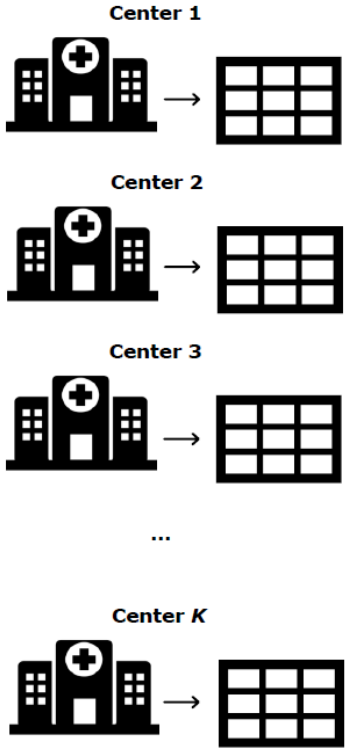
Monitoring for data fabrication

Traditionally:

- On-site visits, source data verification

But presumably more effective:

- Central monitoring of available data
- Allowing for site-by-site comparisons



FDA (2013, p.5): *"Several publications suggest that certain data anomalies (e.g., fraud, including fabrication of data [...]) may be more readily detected by centralized monitoring techniques than by on-site monitoring"*

Aim

Aim: Develop a procedure to detect possible data fabrication in multi-center trials

- Computationally simple
- Generally applicable
- Robust
  - Typos/outliers
  - Variation in number of subjects per center
- 'Semi-automated' application to accumulating trial data
  - without unblinding the data

Goal is to quickly identify centers requiring more detailed assessment, ***not to 'prove' misconduct****.*

For development purposes: Data from the ESPS2 study

Data: The ESPS2 trial

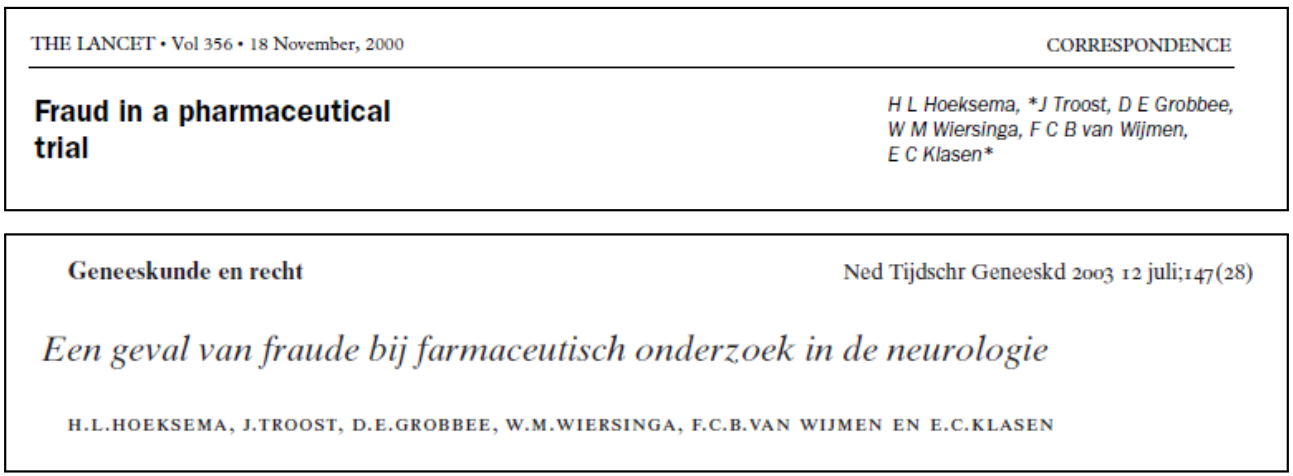
ESPS2: Second European Stroke Prevention Study (1988-1995):

- Double blind, placebo-controlled, randomized trial
- 2x2 factorial design
- Treatments compared (2 years):
  - Placebo
  - Low-dose acetylsalicylic acid
  - Modified-release dipyridamole
  - Combined treatment
- Primary endpoints: Stroke, death
- **13** countries, **60** centers, **6950** subjects

Data: The ESPS2 trial

One fraudulent center ('center 2013'), with **438** subjects:

- Initial suspicions during standard monitoring
- Sample of subjects investigated (directly or via their GP): subjects were unaware of participation in the trial
- Disciplinary action was taken and data was excluded



THE LANCET • Vol 356 • 18 November, 2000 CORRESPONDENCE

Fraud in a pharmaceutical trial

*H L Hoeksema, *J Troost, D E Grobbee, W M Wiersinga, F C B van Wijmen, E C Klasen**

Geneeskunde en recht Ned Tijdschr Geneeskd 2003 12 juli;147(28)

Een geval van fraude bij farmaceutisch onderzoek in de neurologie

H.L.HOEKSEMA, J.TROOST, D.E.GROBBEE, W.M.WIERSINGA, F.C.B.VAN WIJMEN EN E.C.KLASEN

Published central monitoring methods

Fraud detection procedures in the literature:

- Focus: Means, variances, correlations, frequencies, repeated measures, multivariate assessments, digit preference, date checks, etc.
- Methodology:
  - Graphical approaches (e.g. Chernoff faces, star plots, correlation plots)
  - Statistical testing procedures (p-values)

$\rightarrow$ Suitable for our purpose?

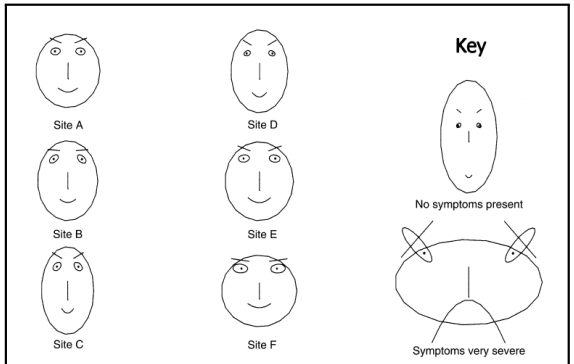


FIGURE 1. Chernoff faces plot of the first 15 questions of the BPRS data split by site.

Source: Taylor et al., 2002

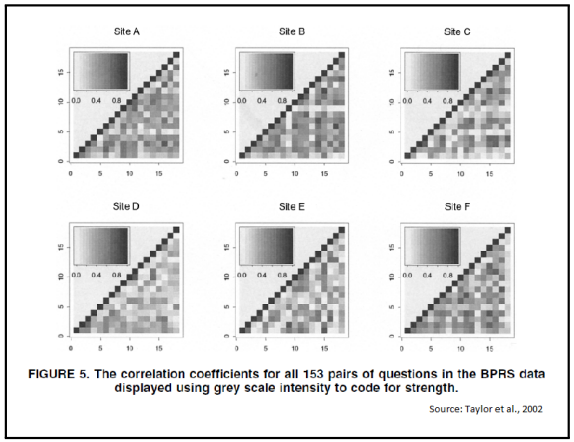


FIGURE 5. The correlation coefficients for all 153 pairs of questions in the BPRS data displayed using grey scale intensity to code for strength.

Source: Taylor et al., 2002

Proposed procedure

- Hypothesized differences between fabricated and real data
  - 7 in total (largely based on published literature)
- Per hypothesized characteristic:
  - Quantification
  - Compare each center to all other centers combined
  - 'Account' for differences in center size by a simple weighting procedure
  - Flag 10% 'most extreme' centers
- If flagged $>$ 2 times $\rightarrow$ More rigorous assessment

For simplicity and general applicability, we only use data collected on baseline

Real vs. fabricated data

Hypothesized characteristics of fabricated data:

1. Differences in 'location' (continuous variables)
2. Lower variability (continuous variables)
3. Deviating pair-wise correlations (continuous variables)
4. Lower proportion of missing values (all variables)
5. Steady recruitment pattern
6. Subject visits taking place during weekends
7. Deviating use of digits (continuous variables)

Quantification of differences

Each center is compared against all other centers combined:

1. Standardized test statistic of Wilcoxon-Mann-Whitney test
2. Ratio of interquartile ranges
3. Difference in Kendall's-tau correlations
4. Proportion difference
5. Deviation from 'perfect' linear pattern
6. Proportion difference
7. Dissimilarity index (bias-corrected)

Note 1: Only centers with >4 subjects are included.
Note 2: If >1 variable, analysis is done for each (results are averaged later).

Weighting

To reduce variability in estimates from small centers, we use the following simple weighting procedure:

- For center i with n_i subjects, the estimate of interest $\hat{\delta}_i$ is downweighted (to 0, in this example) by

$$\hat{\delta}_i^w = \hat{\delta}_i \cdot \frac{n_i}{n_i + m} + 0 \cdot \frac{m}{n_i + m}$$

where m is a tuning parameter (increasing $m \rightarrow$ more shrinkage).

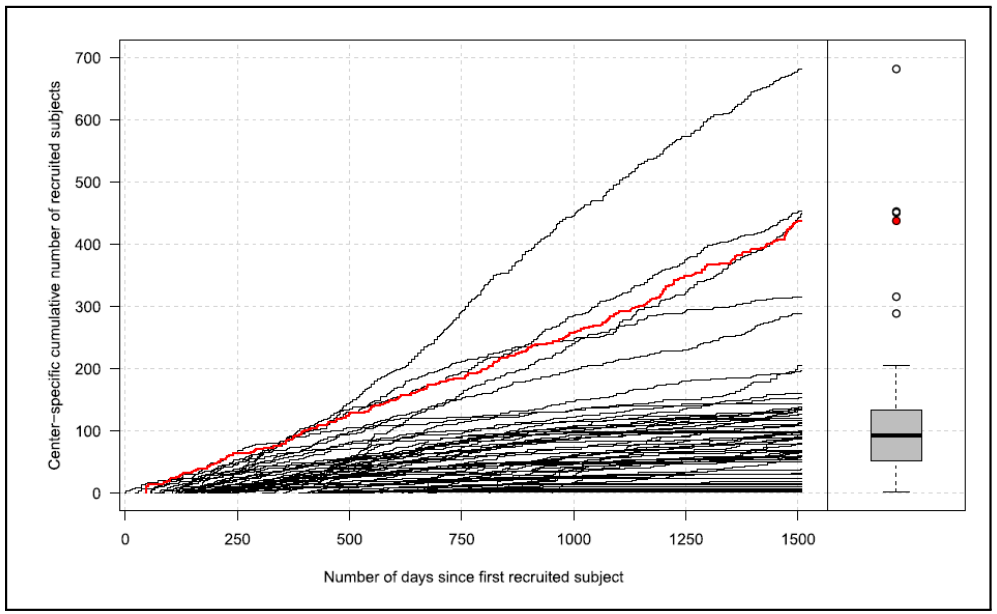
Final steps

- Where applicable (>1 variable), average the weighted estimates
- For each of the 7 analyses, flag 10% most extreme centers
- Count number of flags per center
- If flagged >2 times $\rightarrow$ more detailed assessment

Retrospective application to ESPS2 data

Run procedure on ESPS2 data:

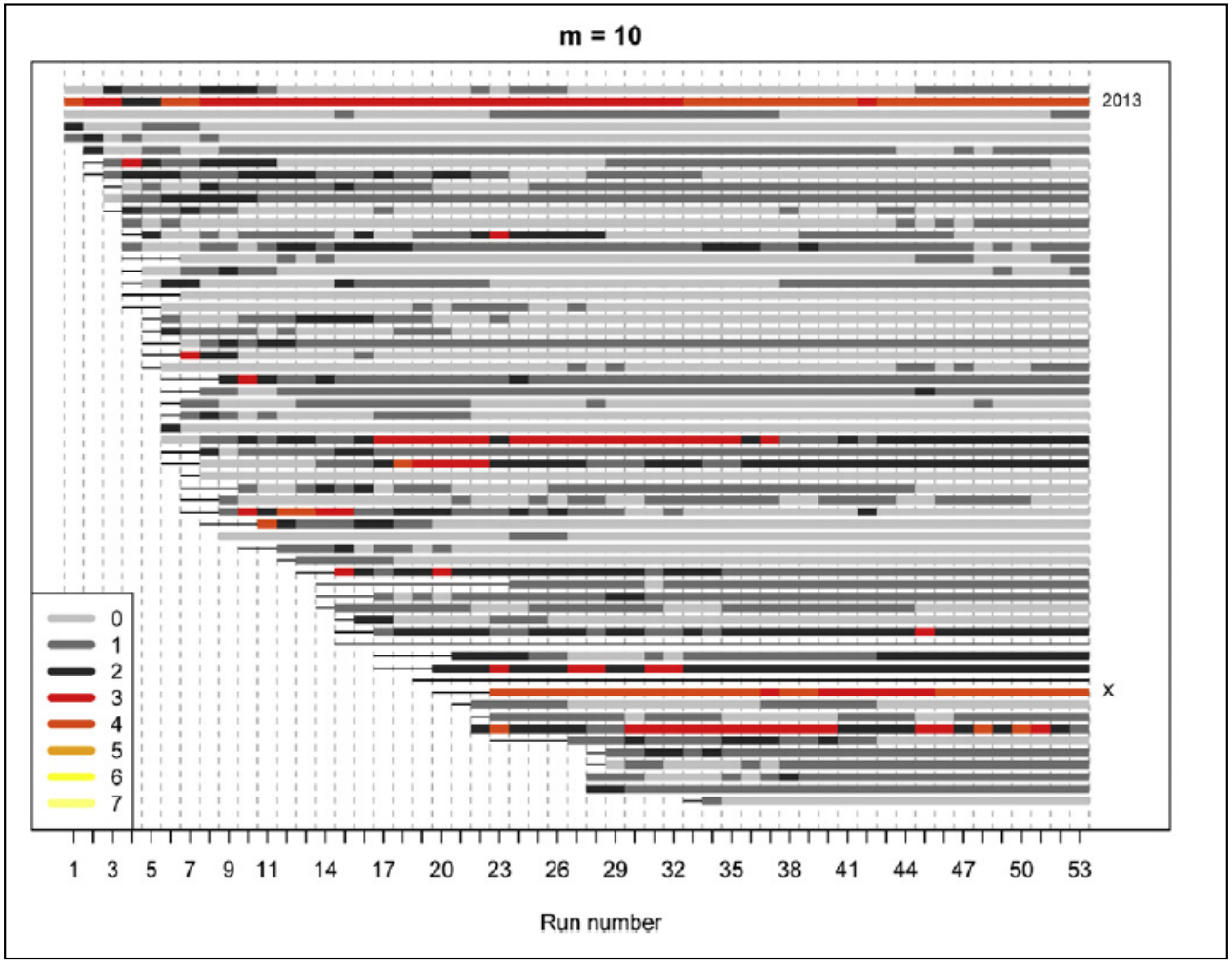
- 53 times, in 4-week intervals
- First run: 5 centers with ≥ 5 subjects
- Last run: Complete dataset
- $m = 0, 5, \mathbf{10}, 20$



Sensitivity analyses (see paper for details):

- Delay start date of center 2013
- Reduce number of subjects
- Combination

Selected results



Selected results

Flag rate for center 2013, per analysis (high to low):

- Lower variability *(flagged in **all** 53 runs)*
- Deviating use of digits *(flagged in **all** 53 runs)*
- Steady recruitment pattern *(flagged in **66%** of runs)*
- Subject visits taking place during weekends *(flagged in **42%** of runs)*
- Lower proportion of missing values *(flagged in **30%** of runs)*
- Differences in 'location' *(flagged in **2%** of runs)*
- Deviating pair-wise correlations *(flagged in **0** runs)*

Selected results

False positive rate:

- On last run: **2%**
- Median over all runs: **4%**

Weighting:

- Best results overall with $m = 5$ or 10
- I.e 'down-weighting' procedure helps

Sensitivity analyses:

- Probability of detection generally remains high
- Except for a few specific combinations (see paper for details)

Discussion/conclusion

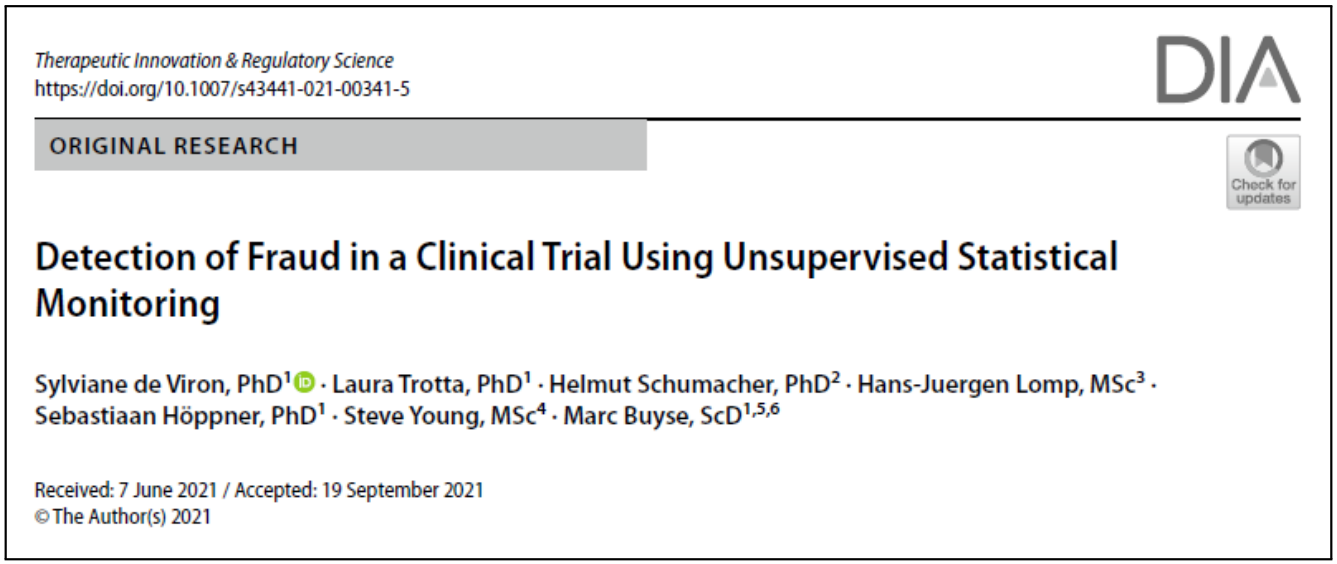
In summary:

- Procedure appears to work well in this dataset
- With methods like this, the fraud could have been spotted very early in the trial

However:

- Hypothesized differences between fabricated and real data may not always be present
- Data fabrication of individual subjects may remain undetected
- Extensions are possible (e.g. utilizing post-baseline data)
- Validation in independent datasets is needed
  - Although a blinded assessment in a recent trial with data fabrication was successful (flagged on 5/7 analyses)
  - Here, results of these analyses helped to identify highly unusual data patterns

Note: Recent publication



Therapeutic Innovation & Regulatory Science
https://doi.org/10.1007/s43441-021-00341-5

DIA

ORIGINAL RESEARCH

Check for updates

Detection of Fraud in a Clinical Trial Using Unsupervised Statistical Monitoring

Sylviane de Viron, PhD[1] · Laura Trotta, PhD[1] · Helmut Schumacher, PhD[2] · Hans-Juergen Lomp, MSc[3] · Sebastiaan Höppner, PhD[1] · Steve Young, MSc[4] · Marc Buyse, ScD[1,5,6]

Received: 7 June 2021 / Accepted: 19 September 2021

"Other approaches have been proposed for the detection of centers with suspicious data. Van den Bor et al. developed a computationally simple statistical procedure to identify the center with known fraud in the ESPS2 trial, using baseline data only. They, too, succeeded in detecting the center 2013. We are currently performing a comparison of the types of discrepancies detected by their statistical procedure, when applied to the totality of the data, as compared to the DQA findings reported here."

Thank you for your attention!

References

- Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. Central statistical monitoring: detecting fraud in clinical trials. Clin Trials 2013;10:225e35.
- Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. Fundamentals of clinical trials. 5th ed. Cham, Heidelberg, New York, Dordrecht, London: Springer; 2015.
- Food and Drug Administration. Guidance for industry: oversight of clinical investigations - a risk-based approach to monitoring. August 2013
- Taylor RN, McEntegart DJ, Stillman EC. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. Ther Innov Regul Sci 2002;36(1):115e25.
- de Viron S, Trotta L, Schumacher H, Lomp H-J, Höppner S, Young S, Buyse M. Detection of fraud in a clinical trial using unsupervised statistical monitoring. Ther Innov Reg Sci 2021.