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

Abstract

Objectives: Central monitoring of multicenter clinical trials becomes an ever more feasible quality assurance tool, in particular for the detection of data fabrication. More widespread application, across both industry sponsored as well as academic clinical trials, requires central monitoring methodologies that are both effective and relatively simple in implementation.

Study Design and Setting: We describe a computationally simple fraud detection procedure intended to be applied repeatedly and (semi-)automatically to accumulating baseline data and to detect data fabrication in multicenter trials as early as possible. The procedure is based on anticipated characteristics of fabricated data. It consists of seven analyses, each of which flags approximately 10% of the centers. Centers that are flagged three or more times are considered ''potentially fraudulent'' and require additional investigation. The procedure is illustrated using empirical trial data with known fraud.

Results: In the illustration data, the fraudulent center is detected in most repeated applications to the accumulating trial data, while keeping the proportion of false-positive results at sufficiently low levels.

Conclusion: The proposed procedure is computationally simple and appears to be effective in detecting center-level data fabrication. However, assessment of the procedure on independent trial data sets with known data fabrication is required.

Keywords: Central monitoring; Fraud detection; Statistical monitoring; Scientific fraud; Investigator misconduct; Risk-based monitoring

1. Introduction

The exact prevalence of fraud or data fabrication (we use the terms ''fraud'' and ''data fabrication'' interchangeably) in clinical trials is difficult to estimate but generally assumed to be low [1–6]. Whether it has a substantial impact on a trial's outcomes depends on the extent and nature of the fraudulent behavior [1,5,7]. 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'' ([5], p.226). Moreover, being aware of recent cases of fraud may deter individuals from participating in the clinical research [8]. Therefore, we agree with Friedman et al., ([9], p.42) who state: ''We condemn all data fabrication. It is important to emphasize that 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.''

Traditionally, quality of trial data is monitored by on-site monitoring and source data verification. However, various authors (e.g., [1,4,5,7,10–12]) have argued that central site-by-site comparisons of digitally available clinical trial data may be more effective than on-site monitoring visits

Funding: Funding was provided by Julius Clinical Ltd. No individuals besides the authors were involved in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Conflicts of interest: None.

Trial information: The methodology proposed in this paper is illustrated using data from the Second European Stroke Prevention Study (ESPS2). ESPS2 was initiated in 1988 and supported by a grant from Boehringer Ingelheim.

* Corresponding author. Tel.: +31-(0)30-656-99-00; fax: +31-(0)30-656-99-90.

E-mail address: r.m.vandenbor@umcutrecht.nl (R.M. van den Bor).

http://dx.doi.org/10.1016/j.jclinepi.2017.03.018

What is new?

Key findings

- We propose a central monitoring procedure that uses a trial for the detection of data fabrication in multicenter clinical trials.
- The procedure appears to be effective when applied to an illustrative empirical data set with known fraud.

What this adds to what was known?

- The proposed strategy is designed to be easily implemented and to be robust against data entry errors and variability in center-specific recruitment rates.
- Therefore, it may improve the probability and/or timing of detecting data fabrication, as it can easily be applied during the data collection phase.
- The R functions that were used are available from the Web Appendix.

What is the implication and what should change now?

- The procedure can be implemented to guide sponsors/clinical research organizations in the detection of potential cases of data fabrication.
- Its performance should be assessed and compared to the performance of existing methodologies using independent empirical data sets, to determine which (combination of) procedure(s) is optimal under which circumstances.

in detecting fraud, for examples of central fraud detection strategies, see [1,2,4,5,11,13–17].

Fraud detection strategies commonly rely on statistical significance tests that are aimed to assess whether a center-specific data pattern deviates (in terms of, e.g., means, variances, digit preference, etc.) from the overall data pattern. The decision to flag a center is then based on the resulting *P*-value. Although this general strategy may yield informative results, the assumption that low *P*-values are indicative for relevant deviations can be problematic, as (1) substantial structural variability is often observed between centers [11] (making a null hypothesis of no difference unrealistic to start with) and (2) the number of observations (i.e., recruited subjects) per center often is highly variable. If so, centers with relatively many observations will be structurally disadvantaged, as was observed by Kirkwood et al. ([4], p.789): "When a large number of data values were examined, even a small difference […] sometimes produced a small *P*-value […]." In addition, if the central monitoring procedure is to be performed on an ongoing basis and/or the number of statistical tests is large, it easily becomes infeasible to assess whether the assumptions of the significance tests are met.

In this paper, we propose an alternative strategy to detect possible fraud in multicenter trials. The strategy shares the aim of detecting deviating data patterns on the center level but uses a weighting procedure, rather than significance tests, to take into account differences in center-specific recruitment numbers. We illustrate the performance of the strategy by applying it repeatedly to accumulating empirical baseline data from a trial with known fraud.

2. Proposed fraud detection strategy

The proposed strategy consists of seven analyses that are based on anticipated characteristics of fabricated data, each of which returns a selection of approximately 10% of the centers that are "most suspicious." The total number of analyses on which a center is flagged then serves as the basis for determining whether a center requires closer inspection. Centers are included if they recruited a minimum of five subjects. All analyses were programmed and performed in R [18]. Details are provided in the following sections.

2.1. Anticipated characteristics of fabricated data

A fraudulent staff member typically does not have access to any trial data besides the data from the subjects recruited by the specific center for which the staff member works. Consequently, we anticipate that, for continuously measured variables, distributions of fabricated observations will be different from the true observations. Fabricated data values may be, on average, too low or too high [1,4,5,19]. Also, data fabrication may become apparent by investigating the spread of the distributions [1,4,5,19,20]. Specifically, we expect variability to be lower in fabricated data because fraudulent investigators either choose to refrain from fabricating extreme values to avoid triggering attention or simply underestimate the variability [20]. Bivariately, deviations may become apparent when comparing pairwise correlations [1,2,4,5,20]. These expectations are assessed in analyses 1, 2, and 3.

In some respects, fabricated data may be expected to be "too perfect." We anticipate that fabricated data will contain relatively few missing values. Also, we expect that the rate by which patients are recruited will be relatively constant over time, as a result of inclusion of either ineligible patients and/or phantom subjects. Analyses 4 and 5 concern these expectations.

Another potential indication of fraud, assessed in analysis 6, is based on the notion that fraudulent investigators may fail to take into account the relative irregularity of subject visits taking place during weekends [1,4,5,7]. Finally, in analysis 7, we compared the distribution of first, second,