

For reprint orders, please contact reprints@future-science.com

Data fraud in clinical trials

Highly publicized cases of fabrication or falsification of data in clinical trials have occurred in recent years and it is likely that there are additional undetected or unreported cases. We review the available evidence on the incidence of data fraud in clinical trials, describe several prominent cases, present information on motivation and contributing factors and discuss cost-effective ways of early detection of data fraud as part of routine central statistical monitoring of data quality. Adoption of these clinical trial monitoring procedures can identify potential data fraud not detected by conventional on-site monitoring and can improve overall data quality.

Keywords: central statistical monitoring • data fabrication and falsification • data fraud • data quality

Honesty and truthfulness are bedrock principles of scientific research. Adherence to these principles is essential both for the progress of science and the public perception of scientific results [1–3]. Deviations from these principles may generally be considered scientific misconduct or fraud, although the US Public Health Service defines research misconduct more narrowly, restricting it to the most egregious practices [4]:

- “Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results;
- Fabrication is making up data or results and recording or reporting them;
- Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record;
- Plagiarism is the appropriation of another person’s ideas, processes, results, or words without giving appropriate credit;

- *Research misconduct does not include honest error or differences of opinion.”*

The last point in the definition is crucial; the key distinction between misconduct or fraud and honest error is intent (fraud is ‘intent to cheat’). For instance, forgetting to report a value is honest error, deliberately not reporting the value is fraud; incorrectly copying a value is honest error, purposely changing the value is fraud. Other governmental bodies in the United States and elsewhere use similar definitions with slight variations. This definition does not explicitly cover other potentially questionable practices in clinical trials that, depending on intent, may be considered scientific misconduct. These practices include selective reporting of results, failure to follow the written protocol, emphasis on secondary rather than primary outcomes, use of improper statistical methods, failure to publish and so on.

In this paper we focus on a specific type of scientific fraud in a specific setting, the deliberate fabrication or falsification of data in clinical trials [5,6]. Although we focus on data fraud, there are many other sources of data errors in clinical trials. These may be more

Stephen L George*¹
& Marc Buyse²

¹Department of Biostatistics & Bioinformatics, Duke University School of Medicine, Durham, NC 27705–3833, USA

²IDDI (International Drug Development Institute) Inc., San Francisco, CA 94109, USA

*Author for correspondence:
Tel.: 1-919-475-1868
stephen.george@duke.edu

common than data fraud, and may have more impact on the conclusions of the trial [7]. Sources of errors can be placed on a continuum, ranging from the honest errors that occur even with the best intentions at one end of the spectrum, to deliberate data fraud at the other end, with bias, misunderstandings, sloppiness and incompetence somewhere in between [8].

In the following sections, we review the available evidence on the incidence of data fraud in clinical trials, describe several prominent cases, present information on motivation and contributing factors and discuss cost-effective ways of early detection of data fraud as part of routine central statistical monitoring of data quality.

How common is data fraud?

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 for many reasons. First of all, in any attempt at direct estimation via a survey of investigators, those who commit fraud are not likely to be forthcoming about having done so. This is a well-known phenomenon to those who study unethical, illegal or any socially unacceptable behavior since respondents have incentives to be evasive. Indirect evidence, based on detected cases, will obviously lead to biased (low) estimates, leading to speculation about a 'tip of the iceberg' phenomenon [6]. The under-reporting of fraud may be even more severe in medical research than in other scientific fields because of the negative consequences such reports have on pharmaceutical companies or other sponsors associated with fraud. There is incentive for sponsors to divulge as little as possible about fraud or misconduct detected in the trials they sponsor. In addition there are definitional problems. In which types of misconduct are we interested? Is it only the narrowly defined serious cases of fabrication, falsification or plagiarism or something broader? And how strictly do we define 'fabricated' data? Suppose an investigator realizes that one specific data point is missing in a case report form, and reports the value observed at the last visit instead of leaving the data point as missing. This, by definition, is data fabrication, in other words, fraud, in contrast to a 'last observation carried forward' approach, which is a legitimate (though often inadequate) statistical technique used in missing data problems. Such instances of data fabrication are likely to be inconsequential, unless they become so common as to dominate actually observed data. Finally, there is a denominator issue. What is the target population? For example, are we interested in the universe of all clinical trials or do we need to distinguish between

small trials conducted at single institutions and large multicenter trials?

Despite these difficulties, there have been numerous attempts over the years to assess the prevalence or incidence of misconduct via surveys, audits and other methods, with conflicting results and conclusions. In a background paper prepared for an ORI conference on research integrity summarizing the literature on research misconduct, the reported estimated incidence of research misconduct ranged over several orders of magnitude [9]. At one end of the scale, based on confirmed cases of fraud (using the narrow PHS definition of fabrication, falsification or plagiarism) a crude estimate of 0.01% (one in 10,000 active scientists) is obtained. A review of routine audits of leukemia trials conducted by the Cancer and Leukemia Group B, one of the multicenter cancer clinical trial groups sponsored by the National Cancer Institute, reported an incidence of fraud of 0.25% [10]. Similarly, a report of audits in the United Kingdom revealed that only 0.40% of the contributed datasets were fraudulent [11]. In the first 10 years of reports from the Office of Research Integrity, there were 136 findings of scientific misconduct (again using the PHS definition); 36 (26%) of these were in clinical trials or other clinical research) [12].

On the other hand, surveys of scientists and medical investigators routinely report rather high levels of misconduct or knowledge of misconduct. These surveys generally fall into one of two types: those that ask about knowledge of misconduct, presumably by others, and those that directly ask about misconduct by those being surveyed. The former provide revealing information about perception of misconduct but do not provide any reliable information about the true incidence; the latter provide direct, but biased, estimates of incidence because of the negative incentives for self-reporting. Examples of the first type include a survey of members of the American Association for the Advancement of Science (AAAS) in which 27% of the scientists reported having encountered some type of misconduct [13]; a survey of research coordinators in which 19% of respondents reported first-hand knowledge of misconduct within the previous year – and that only 70% of these were reported [14]; a study of Norway medical investigators in which 27% of investigators knew of instances of fraud [15]; a survey of members of the International Society of Clinical Biostatistics, in which over 50% of respondents knew of fraudulent reports [16]; a survey of medical institutions in Britain in which more than 50% of respondents knew or suspected misconduct among institutional colleagues [17]; and a survey of New Scientist readers, in which a remarkable 92% knew of or suspected scientific misconduct by colleagues [18].

Examples of the second type of survey include a survey of US academic psychologists, who were asked to self-report ten different questionable research practices, in which high percentages were reported for some practices (e.g., 58% decided to collect more data after ‘peeking’ at the preliminary results) but much less for serious offenses (e.g., 1.7% for falsifying data) [19]; a survey of NIH-funded scientists in which 33% of respondents said they had engaged in one or more of a list of ‘top 10’ questionable behaviors ranging from approximately 16% for ‘changing the design, methodology or results of a study in response to pressure from the funding source’ to 0.3% who admitted to ‘falsification or ‘cooking’ research data’ [20]; and a subsequent similar survey of faculty members of top-tier research institutions in the United States in which nearly 25% of respondents admitted to one or more of the ten most serious types of misconduct during the last 3 years, including 1% who admitted to fabrication or falsification and over 7% who admitted to plagiarism [21]. In a meta-analysis of surveys of questionable research practices from 1987 through 2008, approximately 2% of respondents admitted to data fabrication, falsification or alteration and approximately 34% admitted to other less serious practices [22]. These percentages jumped to 14 and 72%, respectively, in surveys asking about the behavior of colleagues.

Overall, the available evidence is rather unreliable but seems to suggest that the incidence of certain types of misconduct may be quite high but the incidence of data fabrication or falsification is low – but perhaps not as low as might be suggested by the frequency of confirmed and publicized cases. Given the large number of investigators involved, approximately 535,000 US scientists in 2012 [23], even low percentages yield a large number of perpetrators.

Some prominent cases

The history of science is replete with high profile cases of known or suspected scientific misconduct. Indeed, some of the giants of science are not immune from suspicion of questionable practices [24,25], including Claudius Ptolemy, who is suspected of reporting work by others as his own direct observations [26]; Isaac Newton, who may have falsified some data to make them agree more closely to his theories; and Gregor Mendel, who is suspected of some selective reporting of results or even data falsification [27,28]. In these and other examples, there is often no direct proof of fraud, only statistical evidence that the observed results are too close to their theoretically expected value to be compatible with the random play of chance that affects actual experimental data. RA Fisher made that point in a celebrated paper about Mendel’s experiments with

the garden pea, suggesting that Mendel, or one of his assistants, had eliminated outlying observations from his reported results [28]. As we shall discuss later, statistical analysis may similarly point to deviations from expected values in multicenter clinical trials.

Modern examples of scientific fraud are provided in recent papers by Stroebe *et al.* [29], who list a sample of 40 notorious cases and by Sovocool [30], who provides a table of 11 examples of high profile cases. The US Office of Research Integrity, the office charged with investigations of research misconduct by grant supported investigators, provides on-line summaries of the results of their investigations, including the penalties for those found to have committed fraud [31]. There is even a blog (Retraction Watch) listing and discussing retractions of peer-reviewed scientific papers, many of which are the result of scientific misconduct [32]. More than 2000 scientific articles have been retracted over the last 40 years, many in the biomedical sciences. Nine authors have more than 20 retractions apiece [33]. The number of retractions has increased dramatically in recent years and most of these retractions are due to research misconduct, especially data fraud [34]. A brief account of several prominent cases of data fraud in clinical trials illustrates the issues raised by such misconduct.

Roger Poisson

In 1994 it was reported by the leaders of the National Surgical Adjuvant Breast and Bowel Project (NSABP) that a NSABP clinical investigator, Dr. Roger Poisson of St. Luc Hospital in Montréal, Canada had committed scientific fraud by fabrication or falsification of data on several NSABP breast cancer trials from 1977 through 1990 [35,36]. In 1990, a data manager at the NSABP central statistical office noted that some questionable data, especially duplicate operative reports with different dates of surgery, had been submitted by Dr. Poisson. The date changes were such that previously ineligible patients for the trial became eligible. This led to a series of audits at St. Luc that uncovered several instances of altered dates or altered hormone-receptor values. Dr. Poisson subsequently admitted to data falsification and the NSABP reported the matter to the National Cancer Institute (NCI), the sponsor of the NSABP, and to the Office of Scientific Integrity (OSI), the forerunner of the current Office of Research Integrity (ORI). St. Luc had been a major contributor to NSABP trials over the period in question (1511 patients on 22 different trials) but the detailed audits of all patients revealed that only 99 cases (6.3% of all St. Luc patients; 0.3% of all patients entered on the trials) involved any data falsification [6]. Further, the falsification was found to be limited to alterations of

minor eligibility criteria of actual patients. There were no fictitious patients, no violation of the randomization process and no alteration of outcome data. Once entered onto the trial, the patients were treated and followed carefully according to the protocol, and it is now clear that the data fraud in the Poisson case did not compromise either patient safety or affect the overall conclusions of the studies. Unfortunately, due to the importance of the results of the studies for women with breast cancer, the political environment engendered by this importance, the extensive press coverage of the event, the subsequent congressional investigations and the reactions of the sponsor to these events, the perceived importance of the fraud was disproportionate to its actual impact [37].

This case is a cautionary tale for those involved in clinical trials. One of the key strengths of the modern randomized clinical trial, especially large multicenter clinical trials, is the robustness of its conclusions to minor deviations from ideal practice. This is well known to clinical trialists. However, the effect of fraud is likely to be exaggerated in highly publicized cases, especially in high-profile diseases such as breast cancer, and goes far beyond narrow considerations of its scientific impact. Such cases can have a major effect on public perceptions of the results of specific trials and on the value of the clinical trial enterprise in general. These effects are almost certainly more significant than the scientific impact in most cases and, for this reason, cannot be lightly dismissed.

Werner Beswoda

Another case, also in the area of breast cancer trials, but much more serious in nature and scope than the Poisson case, involved Dr. Werner Beswoda, a professor and chair of the Department of Hematology and Oncology at the University of Witwatersrand in Johannesburg, South Africa. In May 1999, at a plenary session presentation at the annual meeting of the American Association of Clinical Oncology (ASCO), Dr. Beswoda presented some striking results of a trial investigating the use of high-dose chemotherapy and stem cell rescue in the treatment of women with high-risk breast cancer [36,38]. The results were notable both because of the positive results (statistically significant differences in both relapse-free survival and overall survival favoring the high-dose chemotherapy group) and because of the contrast with the largely negative results from two other larger but similar studies from the NCI cooperative groups presented in the same session. Beswoda's treatment regimen differed in some key respects from those in the other studies, raising the possibility that these differences were critical in the differing outcomes. The topic of high-dose chemotherapy supported by stem-cell

transplantation in the high-risk breast cancer setting was of very high interest at the time and the presentation of results from these studies had been keenly anticipated. Beswoda's results generated enthusiasm, albeit tinged with skepticism given the results of the other studies.

Shortly after the ASCO meeting, in anticipation of launching a larger and more definitive confirmatory trial based on the Beswoda trial, the NCI commissioned an independent audit team to conduct an on-site audit of records from the trial. The results of that audit were devastating. No medical records of any kind were available for many patients, including over 20% of the patients on the high-dose chemotherapy arm; for patients for whom records could be located there was little evidence that eligibility criteria had been met for nearly two-thirds of the patients; and no evidence could be found that informed consent had been obtained or that the trial had been approved by the appropriate University human research oversight committee [39]. Shortly after the audit results were known, it was announced by the University that Beswoda had been removed from his position and that he had admitted to scientific misconduct.

Robert Fiddes

In the 1990's Dr. Robert Fiddes was the director of the Southern California Research Institute, a for-profit institution, and was the lead clinical investigator for a large number of clinical trials conducted for pharmaceutical company sponsors. Dr. Fiddes was well-known as an investigator who could recruit patients rapidly to clinical trials with a low drop-out rate. Unfortunately, to maintain his highly successful business Dr. Fiddes had been conducting scientific fraud on an impressive scale for over a decade [40]. Ineligible patients were enrolled on trials; fictitious patients were also enrolled; some patients were pressured to enter trials; laboratory data were altered; blood pressure, EKGs and other results were fabricated; blood and urine samples were submitted that did not come from the patient enrolled (in one instance, an employee with proteinuria, a necessary eligibility criterion for some trials, was paid \$25 per urine sample to be submitted as if it were a sample from an actual patient.). Dr. Fiddes was able to maintain his fraud over a long period of time despite audits and other checks until a concerned whistleblower contacted the FDA about the misconduct. After an exhaustive investigation, Dr. Fiddes pled guilty to fraud in 1997, and was sentenced to 15 months in prison [41].

Harry W Snyder Jr & Renee Peugot

In 1994, Dr. Harry W Snyder Jr, a prominent dermatologist and scientist at BioCryst Pharmaceuticals, a

biotech company founded in 1986 in Birmingham, Alabama, was overseeing a clinical trial sponsored by BioCryst at the University of Alabama at Birmingham (UAB), where his wife, Renee Peugot, a registered nurse at the University, was the study coordinator responsible for the day-to-day conduct of the trial. Both had considerable financial interests in BioCryst and a vested interest in the outcome of the trial. The product being tested was BCX-34, a purine nucleoside phosphorylase agent, used as a topical ointment in the treatment of psoriasis and in the treatment of cutaneous T-cell lymphoma (CTCL). Patients with either diagnosis were eligible for the trial. The experimental design was a double-blind placebo-controlled trial with both treatments given to each patient, one treatment (placebo or BCX-34, randomly selected) to lesions on the left side, the other treatment to lesions on the opposite side. In February 1995 a BioCryst press release claimed highly favorable results for BCX-34 for both psoriasis and CTCL patients, particularly noteworthy for CTCL. A subsequent internal re-analysis of the data by the new Medical Director of BioCryst raised serious questions about the initial results and led eventually to a retraction in June 1995 of the results claimed in the earlier press release, now with 'no statistically significant drug effect' noted. Further inquiry and audits by the company and by the FDA led to charges against Snyder and Peugot of falsification of data, including the randomization assignments, to make the results more favorable for BCX-34. Felony convictions followed for both Snyder and Peugot, with prison sentences of 3 years and 2.5 years respectively, payment of financial restitution and permanent debarment by the FDA. The University also had all clinical trials stopped for a period of time while the investigation was underway for failure to properly oversee the trials. [42,43]

Jon Sudbø

In October 2005, *The Lancet* published a paper by Jon Sudbø, a Norwegian physician and researcher, and co-authors on nonsteroidal anti-inflammatory drugs and the risk of oral cancer [44]. The paper reported the analysis of a case-control study of 908 subjects from a population based cohort in Norway. Cases were patients with oral cancer (N = 454) with an equal number of matched controls. Based on detailed statistical analyses, the authors concluded, among other things, that "Long-term use of NSAIDs is associated with a reduced incidence of oral cancer (including in active smokers), but also with an increased risk of death due to cardiovascular disease. These findings highlight the need for a careful risk-benefit analysis when the long-term use of NSAIDs is considered." [44]. However, soon after publication, in January 2006, the editors of *The Lancet* received

correspondence from officials in the hospital where Sudbø worked, that they had uncovered "information that strongly indicates that material published ... has not been based upon data from our national databases, but on manipulated data... it was not manipulation of real data...it was ... complete fabrication." [45]. This was followed quickly by a retraction of the article [46]. A subsequent investigation by an independent commission of inquiry found that all 908 subjects in the *Lancet* paper and all data were fictitious (250 subjects had the same birthday!) and, furthermore, that many of Sudbø's previous publications contained fabricated data, including his doctoral dissertation. Many of these other papers have now also been retracted. Sudbø's dissertation was rescinded and he is no longer allowed to practice medicine or work in medical research.

Yoshitaka Fujii

A letter to the editor of the journal *Anesthesia and Analgesia* in April 2000 called attention to some unusual results in the clinical trials reported in papers published by Yoshitaka Fujii, an anesthesiologist and researcher who had published extensively on his clinical trials involving agents used to treat postoperative nausea and vomiting (PONV) [47]. Specifically, the authors of the letter noted that in 21 papers reporting postoperative headache rates in randomized clinical trials of antiemetic agents, the numbers of headaches was exactly equal in all treatment groups in 13 papers and the numbers differed by at most one in the remaining eight papers. The probability of such a chance occurrence in one trial is quite small; the probability that it occurred in all 21 trials is so small that it strains credulity. The authors, in an ironic understatement, concluded that "...there must be an underlying influence causing such incredibly nice data reported by Fujii et al." [47]. In his brief reply, Fujii failed to address the key issue of how such unusual results might have occurred, noting "...an incidence of headache seems to be identical, but it was true." [48]. Remarkably, despite the suspicions this must have raised and the lack of a satisfactory explanation for the findings, other than this brief exchange of letters, there does not seem to have been any other official follow-up or repercussions at that time. Dr. Fujii continued to publish the results of his clinical trials over the next 12 years.

In 2012, JB Carlisle, a UK anesthesiologist, published an exhaustive analysis of the statistical distributions of variables from 168 randomized clinical trials conducted and published by Fujii over the previous 20 years, an extraordinary number of trials for any investigator [49]. For most of the categorical and continuous variables reported in these papers, the frequency distributions were much less variable than would be expected

by chance alone, echoing and markedly extending the earlier findings on a single variable from a smaller number of studies. Carlisle's cautiously worded conclusion was "Whether the raw data from any of these studies can be analysed, and whether this might provide an innocent explanation of such results...is beyond the scope of this paper." [49]. In his letter replying to this paper, Fujii, as in his previous letter in 2000, again failed to address the key question raised by the statistics (i.e., what is the explanation for these remarkably implausible results?), but stated "...this article by Carlisle can obviously be very damaging to me and I want to answer it seriously, but I am not a statistician. I can only offer a few elements of rebuttal at this point...analyses of data obtained from the experiments were performed by myself and colleagues (co-authors), and this can be proved by them...The only thing I can say is that we performed the tests over years with full honesty and integrity. Additionally, I did not write these articles alone, and some of data were collected by others as well." [50]. But this time the evidence would not be ignored. The editors of 23 journals formally requested a review by the seven institutions in Japan at which Fujii had worked [51]. Shortly thereafter, the Japanese Society of Anesthesiologists (JSA) also began an extensive investigation of 212 papers published by Fujii, including a review of lab records and interviews with other investigators whenever possible. Their findings were astonishing: Out of the 212 papers reviewed, 172 were fraudulent, including 126 'totally fabricated' papers reporting the results of [52]. Meanwhile, as the investigations got underway, Dr. Fujii was no longer involved in research, having been dismissed from the university where he worked for failure to obtain ethical review board approval for his studies.

Why do they do it?

"There's no sense in deliberate falsification, anyhow,' said the Bursar. 'What could anybody gain by it?'; 'It has been done,' said Miss Hillyard, 'frequently. To get the better of an argument. Or out of ambition.'; 'Ambition to be what?' cried Miss Lydgate. 'What satisfaction could one possibly get out of a reputation one knew one didn't deserve? It would be horrible ...I know people do it. But why? They must be mad.'" Dorothy L. Sayers, *Gaudy Night*, 1936.

Most scientists would likely agree with the sentiment expressed in the above quote, suspecting that investigators who commit scientific misconduct either suffer from emotional or mental illness, perhaps with a self-destructive tendency, or have serious character flaws and aberrant personalities. The real reasons in any particular case are likely to be difficult to ascertain. Explanations provided by the perpetrators themselves may not be reliable. In the Poisson case discussed

above, Dr. Poisson wrote a letter explaining why he had falsified data, stating that he was motivated primarily by concern for his patients: "...I believed I understood the reasons behind the study rules, and I felt that the rules were meant to be understood as guidelines and not necessarily followed blindly. My sole concern at all times was the health of my patients. I firmly believed that a patient who was able to enter into an NSABP trial received the best available treatment. For me, it was difficult to tell a woman she was ineligible to receive the best available treatment because she did not meet 1 criterion out of 22, when I knew this criterion had little or no intrinsic oncologic importance" [53].

Souder [54] provides a linguistic comparison of several letters of apology with the official charges of wrongdoing and concludes, perhaps not surprisingly, that "...published acknowledgments of scientific misconduct seem to minimize culpability by means of the strategic use of language...". For example, in a 1999 New York Times article on the Fiddes case it was noted that: "... in interviews with the Government after he agreed to plead guilty, Dr. Fiddes portrayed himself as a man trapped by the dishonesty of others. He maintained that most researchers are forced to cheat because drug companies issue requirements for test subjects that sound good in marketing material, but are impossible to meet in the real world. He said -- with no evidence to back up his claim -- that anyone successful in the business was skirting the rules. ... Dr. Fiddes laid much of the blame for everything that happened on his study coordinators -- again, without providing evidence to support the assertion. While he was the beneficiary of the illegal activity, he maintained that it was the salaried employees working for him who devised the frauds, often without his knowledge. The information provided by Dr. Fiddes has not resulted in any additional investigations" [40].

Presumably the perpetrators of scientific fraud, in common with those who commit fraud in other areas of human activity, do so to obtain some personal advantage (e.g., financial gain – either personally or for research funding, promotion or tenure, awards, prestige, etc.) although in some cases the advantage sought may be difficult to determine. There are also contributing factors besides individual characteristics and much has been written about the reasons that scientists commit scientific misconduct and the conditions that contribute to it [14,20–21,30,55–57]. Davis, *et al.* [56] analyzed 92 cases from the ORI in which the respondent was found to have committed scientific fraud. Starting with 44 possible factors implicated in scientific misconduct, the authors used multidimensional scaling and cluster analysis to define a few clusters of similar factors labelled as personal or professional stressors (e.g., pressure to produce), organizational cli-

mates (e.g., insufficient supervision/mentoring), job insecurities (e.g., competition for position), rationalizations (e.g., lack of control) and personality factors (e.g., laziness). The net effect is to provide a complex picture of the motivations for those who commit scientific fraud. Continued research into the causal factors is important and may help inform more rational preventive measures.

Detection of fraud

Detection of fraud is one aspect of data quality assurance in clinical trials [58]. As part of good clinical practice, trial sponsors are required to monitor the conduct of clinical trials. The aim of monitoring clinical trials is to ensure the patients' well-being, compliance with the approved protocol and regulatory requirements, and data accuracy and completeness [59]. Baigent *et al.* [60] draw a useful distinction between three types of trial monitoring: oversight by trial committees, on-site monitoring and central statistical monitoring, and argue that the three types of monitoring are useful in their own right to guarantee the quality of the trial data and the validity of the trial results. Oversight by trial committees is especially useful to prevent or detect errors in the trial design and interpretation of the results. On-site monitoring is especially useful to prevent or detect procedural errors in the trial conduct at participating centers (e.g., whether informed consents have been signed by all patients or legally acceptable representatives). Statistical monitoring is especially useful to detect data errors, whether due to faulty equipment, sloppiness, incompetence or fraud.

A survey of current monitoring practices reveals that the vast majority of trials are monitored primarily through on-site visits with source data verification, which consists of comparing information recorded in the trial's case report form with the corresponding source documents [61]. While there is general agreement that some on-site monitoring is necessary, the role of source data verification, especially extensive source data verification, is increasingly being questioned [62]. Source data verification detects discrepancies due to transcription errors from source documentation to the case report form, but not errors present in the source documents. Source data verification may be useful to guarantee that the primary outcome of the trial and some key safety parameters have been accurately captured, but full (100%) verification of all source data is particularly cost-ineffective [63–65]. Recent guidance documents from the US Food and Drug Administration and the European Medicines Agency strongly favor the use of 'quality by design' and risk-based monitoring approaches, instead of traditional monitoring

techniques that have proven to be costly as well ineffective [66,67]. In particular, exhaustive source data verification may be replaced by targeted data audits when indicated.

It seems somewhat paradoxical that statistical theory, which is so central to the design and analysis of clinical trials, has not yet been put to use to help optimize monitoring activities, even though the potential of statistics to uncover fraud in multicenter trials has received attention for more than a decade [68,69]. The recent regulatory guidance documents [66,67] have spurred much interest in using central statistical monitoring as a tool to detect fraud and, more generally, any abnormal pattern in the data that could help focus monitoring activities on centers where they appear to be most needed [70,71]. The second European Stroke Prevention Study (ESPS2) provides a impressive example of the effectiveness of statistical monitoring as compared with on-site visits for the detection of abnormal data patterns [72]. ESPS2 was a randomized trial of aspirin and dipyridamole in patients with transient ischemic attack or stroke. The study enrolled 6602 patients at 59 centers, plus 438 patients at a center that eventually had to be excluded from all analyses. *"Fraud or misconduct at the center concerned was considered a possibility early in the recruitment. Despite intensive monitoring this could not be proven one way or the other and external audit was brought in. The audit also failed to establish guilt or innocence..."* [72]. In the end, the center was excluded on the grounds that the distribution of dipyridamole and aspirin plasma concentrations differed significantly in the suspect center as compared with all other centers, and was incompatible with the drug administration required by the study protocol. This case exemplifies a situation in which even the most careful on-site review cannot uncover unusual data patterns that are readily detected even by simple statistical methods to compare distributions of continuous variables [68]. More sophisticated statistical methods can detect less obvious data patterns.

Principles of statistical monitoring

Statistical monitoring of clinical trials uses a few basic procedures based on the nature of data collected. First, statistical monitoring relies on the highly structured nature of clinical data, since the same protocol is implemented identically in all participating centers, where data are collected using the same case report form, whether on paper or electronically. Abnormal trends and patterns in the data can be detected by comparing the distribution of some or all variables in each center against all other centers [73–76]. Similar comparisons can also be made between other units of analysis, if the structure of the trial warrants it. When

the trial is randomized, the treatment group allocated by randomization provides another design feature that allows for specific statistical tests to be performed [68]. Indeed, baseline variables are not expected to differ between the randomized groups (but through the play of chance), while outcome variables are expected to differ about equally in all centers (but through the play of chance).

A second tenet of statistical monitoring is that even when simple comparisons indicate no major differences in the data of all centers, a more in-depth investigation of the complex data structure typical of clinical trials can be informative. The multivariate structure or time dependence of the variables can provide the basis for sensitive tests of data quality. Fabricated or falsified data, even if plausible univariately, are likely to exhibit abnormal multivariate patterns that are hard to mimic and therefore easy to detect statistically [5,69]. The frequency of ‘data collisions,’ in other words, identical values for one or several variables for different patients, is a sensitive indicator of situations where data have simply been cut and pasted from one patient to another. Similarly, variables that are repeatedly measured over time can be statistically scrutinized for ‘data propagation.’ In addition, humans are poor random number generators, and are generally forgetful of natural constraints in the data. Tests on randomness can be used to detect invented data. Benford’s law on the distribution of the first digits, or tests for digit preference, can raise red flags [77,78]. Tests on dates can also be used to detect anomalies in the distribution of days (e.g., a high proportion of visits during weekends may reveal data fabrication) [5].

Which of the cases of fraud discussed above could have been detected by statistical methods? Although the answer to this question remains conjectural for the examples cited, the anomalies in Dr. Poisson’s center were detected by the NSABP statistical office, although not as part of a routine central statistical monitoring program as discussed in this paper. The Sudbø and Fujii cases would almost surely have been detected, since the data reported by these two investigators contained gross aberrations that in retrospect appear too gross to have remained unnoticed upon close scrutiny. The Fiddes case could arguably have been detected since the trials involved were multicentric, but the Snyder/Peugot and Bezwoda cases would have been far more difficult to detect statistically since these trials were carried out at single institutions.

Implementations of statistical monitoring

Different implementations of statistical monitoring have been proposed in the literature [73–76,79–80]. The most popular approach is based on ‘key risk indica-

tors,’ which are clinical data variables identified as important, and monitored throughout the trial against pre-specified thresholds [73,79]. A site that exceeds the threshold for a key risk indicator is flagged for further scrutiny. For instance, protocol violations could constitute a key risk indicator. Sites could be flagged if they experienced protocol violations in more than, say, 10% of their patients. Although key risk indicators are quite useful as part of routine clinical trial monitoring, their potential for data fraud detection is quite limited. A more sophisticated approach was developed specifically to detect fraud in cardiovascular trials [80]. Predictive models were built using the database from a multicenter trial in which data from 9 out of 109 centers had been documented to be fabricated. The predictive models used a few key variables such as the systolic and diastolic blood pressure to generate risk scores for each center. The risk scores had the ability to discriminate well between centers with and without fabricated data, and were validated using an independent multicenter trial database that contained no data fabrication. Whether simple models based on a few variables can be generally useful requires further validation [80]. Besides, the variables entering these predictive models would have to depend on the disease and treatment studied. In cancer trials, for instance, blood pressure is not a relevant indicator of risk; it is in fact often not measured at all.

A full statistical approach to data monitoring consists of performing as many statistical tests as possible on as many clinical data variables as possible: tests for proportions, means, global variances, within-patient variances, event counts, distributions of categorical variables, proportion of week days, outliers, missing values, correlations between several variables and so forth [74–76]. The central idea is to compare the data of each center with the data of all other centers, which requires no distributional assumptions and can therefore be largely automated [76]. Extensive testing of all variables in a clinical trial raises challenging issues, including control of multiplicity and avoidance of false-positive signals, but there are statistical ways of addressing these issues [81]. Graphical displays can help spot centers with data anomalies [74,75]. Alternatively, the statistical tests can generate a high-dimensional matrix of p-values, with centers as rows and tests as columns, analogous to the gene expression matrix of a microarray experiment. Principal component analysis can be used to reduce the dimensionality of the matrix and help detect outlying centers [76]. A ‘data inconsistency score’ (DIS) can also be calculated for each center, for example, as the mean of minus the log p-values of all statistical tests performed. Such a score ranges from 0 to 10 (or in very rare instances larger) and has there-

fore a similar interpretation to the Richter scale for the amplitude of seismic waves. Centers with a score close to 0 have data that are compatible with data from other centers, while centers with high scores (say, 3 or above) have data that are so inconsistent with data from other centers that the observed differences cannot be attributed to chance alone. Centers with extreme data inconsistency scores are worthy of further investigation, with the aim of explaining the differences, retraining the site personnel if required, or – in the worst case scenario – to uncover a fraud that would otherwise remain undetected. Note that centers can have inconsistent data if they treat different patient populations, so high data inconsistency scores do not automatically imply that a remedial action is warranted. Also note that centers can have inconsistent data if their data are of much better quality than on average (say, for instance, if they have far fewer missing data than overall). Hence high data inconsistency scores are a statistical finding with no implied value judgment and should not be interpreted as a ‘data quality’ index.

An example of fraud

Statistical monitoring of clinical trial data is likely to pick up many cases of fraud simply because of unusual patterns in the data. As an example, Figure 1 shows a

‘bubble plot’ produced using central statistical monitoring of a completed randomized clinical trial involving more than 4500 patients treated in 160 clinical centers (details omitted to preserve anonymity). In the bubble plot, each center is represented by a bubble, the size of which is proportional to the center size. The horizontal position of the bubble is also proportional to the center size, while the vertical position of the bubble represents the data inconsistency score of the center. Bubbles falling above the horizontal line labeled ‘FDR = 3%’ correspond to centers with extreme data inconsistency scores, indicating that the data collected in these centers differ statistically from the data collected in all other centers. The false discovery rate (FDR) above the horizontal line is less than 3%, in other words, there is less than 3% chance that any of the sites above the line was identified as having inconsistent data just by the play of chance.

A fraud was known to have occurred at center X, where 97 patients had been treated. An on-site audit had revealed that in center X some patients had not been provided with quality of life questionnaires, which had instead been completed by site personnel [76]. Interestingly, statistical analysis of the data suggested that the quality of life data were equally suspicious in center Y, yet on-site visits at that center had not uncovered

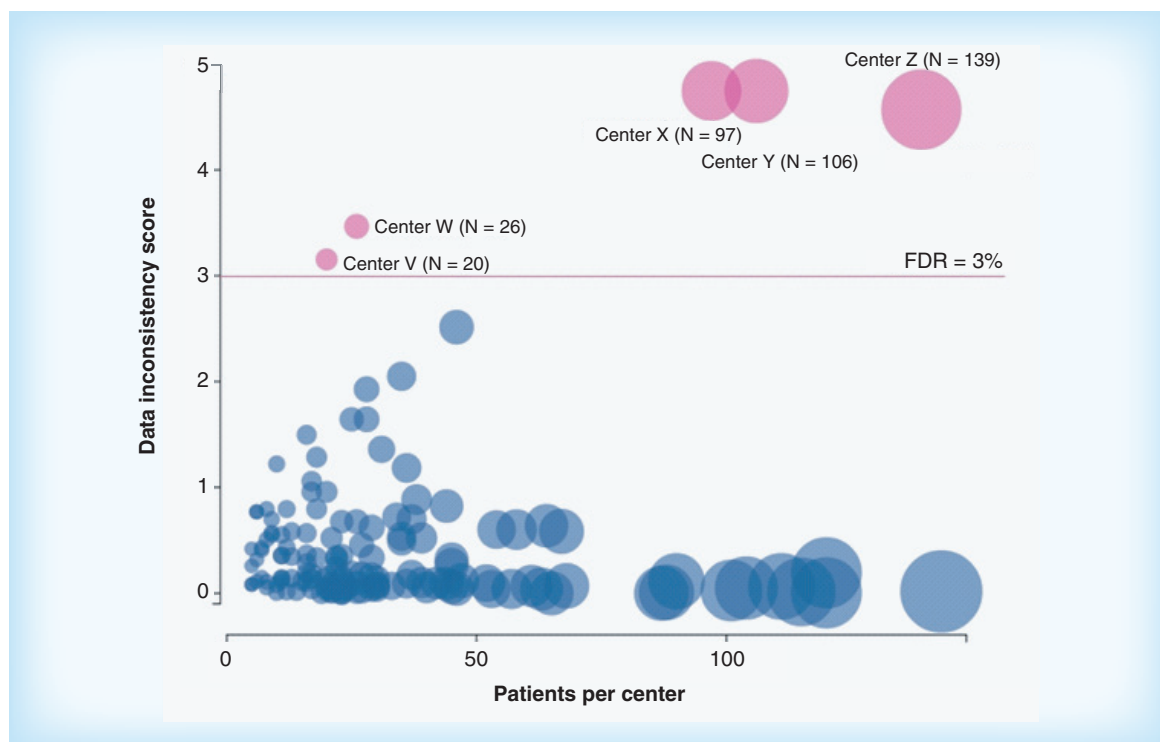


Figure 1. Bubble plot showing the data inconsistency score as a function of the center size in a randomized clinical trial. The data from the centers above the horizontal line (labeled V, W, X, Y and Z) differ statistically from the data in all other centers. Fraud was confirmed to have occurred in center X (more details on the example can be found in [76]).

FDR: False discovery rate; N: Number of patients per center.

any problem at that center. This example illustrates the effectiveness of statistical analysis to reveal fabricated data, which tend to differ in several subtle but detectable ways from actually observed data [68].

An example of a technical problem

Statistical monitoring of clinical trial data may pick up any unusual pattern in the data, whether due to fraud, tampering, sloppiness, incompetence, misunderstanding, technical problems or any other cause. As an example, Figure 2 shows the distribution of mean body temperatures in an on-going randomized clinical trial involving more than 16,000 patients treated in 218 clinical centers (details omitted to preserve anonymity).

The mean body temperatures in the 12 centers of country A were within the allowed range and were not flagged as suspicious by data management checks or during on-site monitoring visits. Yet when the totality of the data was submitted to statistical monitoring checks, the centers in country A were clearly identified as inconsistent with all other centers. After further investigation, it was found that a single lot of thermometers was miscalibrated in this country, causing a downward shift in the temperatures too small to be detected in a single measurement, but statistically detectable in a large number of them [81].

Conclusion

There is no reliable evidence that data fraud, the deliberate fabrication or falsification of data, is a common occurrence in clinical trials. Moreover, in multicenter clinical trials, fraud perpetrated by a single investigator

or at a single site is very unlikely to affect the scientific conclusions of the trial. However, whatever the true incidence of data fraud in clinical trials, high-profile cases provide sobering evidence that it does occur regularly. When fraud is detected after the results have been announced, the negative impact on the perception of the results of the trial in question as well as on the public perception of the clinical trial enterprise itself can be profound. In addition, inadequate training of site personnel, misunderstandings and sloppiness can also result in incorrect data being contributed by some sites in multicenter trials. For these reasons it is important to have procedures in place as early as possible to identify patterns that might indicate data issues. In this paper we describe some cost-effective central statistical monitoring procedures that, when included as part of an overall data quality assurance program, are likely to detect instances of data fraud as well as other data problems at an early, treatable time point during the trial. Knowledge that such procedures are in place may also act as a deterrent to fraud in the first place. With modern methods of data collection and modern statistical computing capabilities there are few impediments to implementing such a system to supplement or reduce the standard on-site monitoring and risk-based procedures.

Future perspective

Procedures for ensuring the quality of data in clinical trials, including but not limited to detection and treatment of data fraud, will continue to expand and mature. The development and use of cost-effective computer-based automated approaches for quality control and quality assurance will become even more

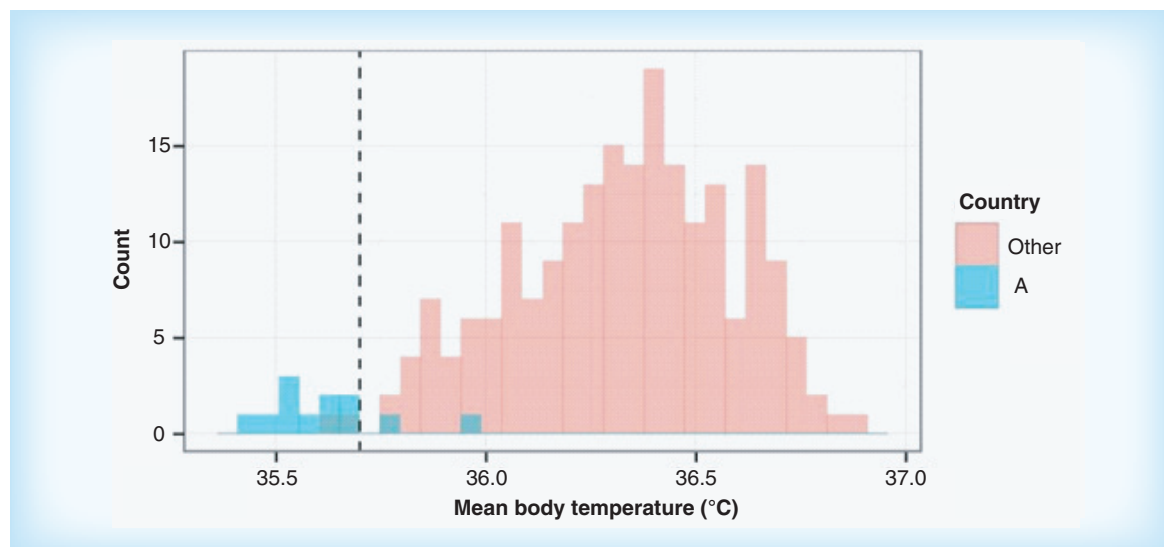


Figure 2. Distribution of mean body temperatures (in °C) for centers in country A and for all other centers in a randomized clinical trial. The data from 10 of the 12 centers in country A differ statistically from the data in all other centers. The difference was found to be due to a single lot of miscalibrated thermometers. Adapted with permission from [81] © John Wiley and Sons (2014).

important as the costs of clinical trials continue to rise. Central statistical monitoring techniques of the type discussed in this paper will be more heavily used, since they have much to offer in ensuring the data integrity of clinical trials. While these techniques will be used primarily by trial sponsors to suggest remedial actions during the trial, they will also prove valuable to provide regulatory agencies with an overall assessment of the data quality in trials submitted as part of a marketing authorization. Similarly, a statistical assessment of data quality may prove quite useful during peer review. It seems likely and desirable that in the future, journal editors will increasingly request access to the source data upon which claims are made, and that peer review will routinely include a statistical assessment of data quality.

Financial & competing interests disclosure

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Number P01CA142538. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. SL George declares no conflict of interest. M Buyse holds stock in IDDI, Inc. and CluePoints, Inc., a company that offers central statistical monitoring services. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Data fraud, the deliberate fabrication or falsification of data, is a serious type of scientific misconduct in clinical trials.
- The incidence or prevalence of data fraud in clinical trials is generally assumed to be quite low but, by its nature, this incidence is difficult to estimate. In addition to the highly publicized cases of extreme data fabrication or falsification, many less extreme instances may remain undetected by conventional on-site monitoring.
- In multicenter clinical trials the effect on the primary conclusions of the trial of isolated instances of data fraud by a single investigator or site is almost certainly negligible.
- Serious harmful effects of data fraud in clinical trials include potential harm to patients, negative publicity and the resulting negative public perception of the results of the trial in question and broader damage to the clinical trial enterprise itself.
- It is important to have procedures in place to identify potential cases of fraud as early as possible during the conduct of the trial so such cases can be investigated and resolved before damage has occurred.
- There are cost-effective ways to incorporate detection of potential fraud as part of routine central statistical monitoring of data quality.
- Adoption of a comprehensive central statistical monitoring plan will minimize the possibility of damage to trial integrity should a case of data fraud occur and will likely have a deterrent effect as well.

References

- 1 Institute of Medicine, National Academy of Sciences, National Academy of Engineering. *On Being a Scientist: A Guide to Responsible Conduct in Research (3rd Edition)*. The National Academies Press, Washington, DC, USA (2009).
- 2 National Research Council. *The Responsible Conduct of Research in the Health Sciences*. The National Academies Press, Washington, DC, USA (1989).
- 3 Resnik DB. *The Ethics of Science: An Introduction*. Routledge, London (1998).
- 4 Federal Register. Public health service policies on research misconduct final rule. 42 CFR part 93. <http://www.ecfr.gov/cgi-bin/text-idx?SID=0b07ed68cf889962cae6c2b45d89150b&node=pt42.1.93&rgn=div5> (2005).
- 5 Buyse M, Evans SJW. Fraud in clinical trials. In: *Encyclopedia of Biostatistics*. Armitage P, Colton T (Eds). John Wiley, Chichester 2023-2031 (2005).
- 6 George SL. Perspectives on scientific misconduct and fraud in clinical trials. *Chance* 10(4), 3–5 (1997).
- 7 Lederberg J. Sloppy research extracts a greater toll than misconduct. *Scientist* 20, 13 (1995).
- 8 Demets DL. Distinctions between fraud, bias, errors, misunderstanding, and incompetence. *Controlled Clinical Trials* 18(6), 637–650 (1997).
- 9 Steneck NH. Assessing the integrity of publicly funded research: a background report for the November 2000 ORI Research Conference on Research Integrity. In: ORI Research Conference on Research Integrity. Bethesda, MD, USA (2000).
- 10 Weiss RB, Vogelzang NJ, Peterson BA *et al*. A successful system of scientific data audits for. A report from the Cancer and Leukemia Group B. *JAMA* 270(4), 459–464 (1993).
- 11 Hone J. Combating fraud and misconduct in medical research. *Scrip Magazine* 14(March), 14–15 (1993).
- 12 Reynolds SM. ORI findings of scientific misconduct in clinical trials and publicly funded research, 1992–2002. *Clin. Trials* 1(6), 509–516 (2004).
- 13 Hamilton D. In the trenches, doubts about scientific integrity. *Science* 255(5052), 1636 (1992).

- 14 Habermann B, Broome M, Pryor ER, Ziner KW. Research coordinators' experiences with scientific misconduct and research integrity. *Nurs. Res.* 59(1), 51–57 (2010).
- 15 Jacobsen G, Hals A. Medical investigators' views about ethics and fraud in medical research. *J. R. Coll. Physicians Lond.* 29(5), 405–409 (1995).
- 16 Ranstam J, Buysse M, George SL *et al.* Fraud in medical research: an international survey of biostatisticians. ISCB Subcommittee on Fraud. *Controlled Clinical Trials* 21(5), 415–427 (2000).
- 17 Farthing MJG. Research misconduct: diagnosis, treatment and prevention. *Br. J. Surg.* 87(12), 1605–1609 (2000).
- 18 Howard E. Science misconduct and due process: a case of process due. *Hastings L. J.* 45, 309 (1993).
- 19 John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23(5), 524–532 (2012).
- 20 Martinson BC, Anderson MS, De Vries R. Scientists behaving badly. *Nature* 435(7043), 737–738 (2005).
- 21 Martinson BC, Crain AL, Anderson MS, De Vries R. Institutions' expectations for researchers' self-funding, federal grant holding, and private industry involvement: Manifold drivers of self-interest and researcher behavior. *Acad. Med.* 84(11), 1491–1499 (2009).
- 22 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4(5), e5738 (2009).
- 23 Sargent Jr JF. The US Science and Engineering Workforce: Recent, Current, and Projected Employment, Wages, and Unemployment (2014).
- 24 Broad W, Wade N. *Betrayers of the Truth: Fraud and Deceit in Science.* Simon and Schuster, New York (1982).
- 25 Judson HF. *The Great Betrayal: Fraud in Science.* Harcourt, Orlando. (2004).
- 26 Newton RR. *The Crime of Claudius Ptolemy.* Johns Hopkins University Press, Baltimore. (1977).
- 27 Galton D. Did Mendel falsify his data? *Q. J. M.* 105(2), 215–216 (2012).
- 28 Fisher RA. Has Mendel's work been rediscovered? *Annals of Science* 1, 115–137 (1936).
- 29 Stroebe W, Postmes T, Spears R. Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science* 7(6), 670–688 (2012).
- 30 Sovacool BK. Exploring scientific misconduct: isolated individuals, impure institutions, or an inevitable idiom of modern science? *Bioethical Inquiry* 5(4), 271–282 (2008).
- 31 Office of Research Integrity. Case Summaries. http://ori.hhs.gov/case_summary
- 32 Retraction Watch. <http://retractionwatch.com/>
- 33 Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS ONE* 8(7), e68397 (2013).
- 34 Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *PNAS* 109(42), 17028–17033 (2012).
- 35 Fisher B, Redmond CK. Fraud in breast-cancer trials [letter]. *N. Engl. J. Med.* 330(20), 1458–1460 (1994).
- 36 Weir C, Murray G. Fraud in clinical trials. *Significance* 8(4), 164–168 (2011).
- 37 Peto R, Collins R, Sackett D *et al.* The trials of Dr. Bernard Fisher: a European perspective on an American episode. *Controlled Clinical Trials* 18(1), 1–13 (1997).
- 38 Horton R. After Bezwoda. *Lancet* 355(9208), 942–943 (2000).
- 39 Weiss RB, Rifkin RM, Stewart FM *et al.* High-dose chemotherapy for high-risk primary breast cancer: an on-site review of the Bezwoda study. *Lancet* 355(9208), 999–1003 (2000).
- 40 Eichenwald K, Kolata G. A doctor's drug studies turn into fraud. *N. Y. Times Web* A1, A16–A11, A16 (1999).
- 41 Swaminathan V, Avery M. FDA enforcement of criminal liability for clinical investigator fraud. *Hastings Science and Technology Law Journal* 4, 325–356 (2012).
- 42 Birch DM, Cohen G. How a cancer trial ended in betrayal. *Baltimore Sun.* <http://www.baltimoresun.com/bal-te.research24jun24-story.html#page=1> (2001).
- 43 Grant B. Biotech's Baddies. *Scientist* 23(4), 48 (2009).
- 44 Sudbø J, Lee JJ, Lippman SM *et al.* RETRACTED: non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *Lancet* 366(9494), 1359–1366.
- 45 Horton R. Expression of concern: non-steroidal anti-inflammatory drugs and the risk of oral cancer. *Lancet* 367(9506), 196 (2006).
- 46 Horton R. Retraction – Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *Lancet* 367(9508), 382 (2006).
- 47 Kranke P, Apfel CC, Roewer N. Reported data on granisetron and postoperative nausea and vomiting by Fujii *et al.* are incredibly nice! [letter]. *Anesth. Analg.* 90(4), 1004 (2000).
- 48 Fujii Y. Reported data on granisetron and postoperative nausea and vomiting by Fujii *et al.* are incredibly nice! *Anesth. Analg.* 90(4), 1004 (2000).
- 49 Carlisle J. The analysis of 168 randomised controlled trials to test data integrity. *Anaesthesia* 67(5), 521–537 (2012).
- 50 Fujii Y. The analysis of 168 randomised controlled trials to test data integrity. *Anaesthesia* 67(6), 669–670 (2012).
- 51 Miller D. Retraction of articles written by Dr. Yoshitaka Fujii. *Can. J. Anesth.* 59(12), 1081–1088 (2012).
- 52 Normile D. A new record for retractions? *Science Insider* (2012).
- 53 Poisson R. Fraud in breast-cancer trials [letter]. *N. Engl. J. Med.* 330(20), 1460 (1994).
- 54 Souder L. A rhetorical analysis of apologies for scientific misconduct: do they really mean it? *Sci. Eng. Ethics* 16(1), 175–184 (2010).
- 55 Garshick MK, Kimball AB. Research integrity in the modern era: current gaps in our knowledge and thinking. *Clinical Investigation* 4(3), 199–200 (2014).

- 56 Davis MS, Riske-Morris M, Diaz SR. Causal factors implicated in research misconduct: evidence from ORI case files. *Sci. Eng. Ethics* 13(4), 395–414 (2007).
- 57 Adams D, Pimple KD. Research misconduct and crime lessons from criminal science on preventing misconduct and promoting integrity. *Account. Res.* 12(3), 225–240 (2005).
- 58 Knatterud GL, Rockhold FW, George SL *et al.* Guidelines for quality assurance procedures for multicenter trials: A position paper. *Controlled Clinical Trials* 19(5), 477–493 (1998).
- 59 International Conference on Harmonisation Working Group: ICH harmonised tripartite guideline: guideline for good clinical practice E6 (R1). In: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. (1996).
- 60 Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clin. Trials* 5(1), 49–55 (2008).
- 61 Morrison BW, Cochran CJ, White JG *et al.* Monitoring the quality of conduct of clinical trials: a survey of current practices. *Clin. Trials* 8(3), 342–349 (2011).
- 62 Smith CT, Stocken DD, Dunn J *et al.* The value of source data verification in a cancer clinical trial. *PLoS ONE* 7(12), e51623 (2012).
- 63 Eisenstein EL, Collins R, Cracknell BS *et al.* Sensible approaches for reducing clinical trial costs. *Clin. Trials* 5(1), 75–84 (2008).
- 64 Reith C, Landray M, Devereaux P *et al.* Randomized clinical trials – removing unnecessary obstacles. *N. Engl. J. Med.* 369(11), 1061–1065 (2013).
- 65 Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. *Lancet* 366(9480), 172–174 (2005).
- 66 European Medicines Agency. Reflection Paper on Risk Based Quality Management in clinical trials. EMA/INS/GCP/394194/2011 (2011).
- 67 Food and Drug Administration. Guidance for Industry: Oversight of Clinical Investigations – a Risk-Based Approach to Monitoring. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM269919.pdf
- 68 Buyse M, George SL, Evans S *et al.* The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine* 18(24), 3435–3451 (1999)
- 69 Evans S. Statistical aspects of the detection of fraud. In: *Fraud and Misconduct in Medical Research (3rd Edition)*. Lock S, Wells F, Farthing M (Eds). BMJ Publishing Group, London 186–204 (2001).
- 70 Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. *Clin. Trials* 9(2), 257–264 (2012).
- 71 Buyse M. Centralized statistical monitoring as a way to improve the quality of clinical data. *Applied Clinical Trials*, 24 March 2014.
- 72 European Stroke Prevention Study 2 Group. European Stroke Prevention Study 2: efficacy and safety data. *J. Neurol. Sci.* 151(Suppl), S1–S77 (1997).
- 73 Edwards P, Shakur H, Barnetson L, Prieto D, Evans S, Roberts I. Central and statistical data monitoring in the Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage (CRASH-2) trial. *Clin. Trials* 1(3) 336–343 (2013).
- 74 Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. *Clin. Trials* 10(5), 783–806 (2013).
- 75 Lindblad AS, Manukyan Z, Purohit-Sheth T *et al.* Central site monitoring: Results from a test of accuracy in identifying trials and sites failing Food and Drug Administration inspection. *Clin. Trials* 11(2), 205–217 (2013).
- 76 Venet D, Doffagne E, Burzykowski T *et al.* A statistical approach to central monitoring of data quality in clinical trials. *Clin. Trials* 9(6), 705–713 (2012).
- 77 Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *Br. Med. J.* 331(7511), 267–270 (2005).
- 78 Hill TP. A statistical derivation of the significant-digit law. *Stat. Sci.* 10(4), 354–363 (1995).
- 79 Valdés-Márquez E, Hopewell CJ, Landray M, Armitage J. A key risk indicator approach to central statistical monitoring in multicentre clinical trials: method development in the context of an ongoing large-scale randomized trial. *Trials* 12(Suppl 1), A135 (2011).
- 80 Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. Central statistical monitoring: detecting fraud in clinical trials. *Clin. Trials* 10(2), 225–235 (2013).
- 81 Desmet L, Venet D, Doffagne E *et al.* Linear mixed-effects models for central statistical monitoring of multicenter clinical trials. *Statistics in Medicine* 33(30), 5265–5279 (2014).