

Fraud in clinical trials

Detecting it and preventing it

Fraud is surprisingly hard to conceal. If you make up or alter the data in a clinical trial you will leave a trail that a good statistician can follow. **Christopher Weir** and **Gordon Murray** tell would-be detectives where the clues will be.

If a man defrauds you one time, he is a rascal; if he does it twice, you are a fool.

(Author unknown)

Clinical trials are important. Much hangs on their results – for the patients, new and better treatments; for the clinicians, prestige, promotion, pay. A trial that leads to negative results gets little thanks or credit within the medical profession – name-making academic papers do not follow. Given all that, the incentive to fabricate or to falsify results in a clinical trial is clearly large.

Which makes it important that in setting the scene we emphasise that, based on the available evidence,

fraud in clinical research is uncommon. Of the 22 cases from clinical trials recorded in the Committee for Publication Ethics database between 1997 and 2011, most relate to ethical review and only three to investigation of potential data fabrication. Other estimates suggest that fraud occurs in rather less than 1% of publicly and commercially supported medical research. This may of course be an underestimate; by its nature undetected fraud is unrecorded fraud. Box 1 summarises two widely publicised fraud cases.

Case 1 would have been best addressed through the application of rigorous institutional approval and ethical review processes and may be less likely to occur now that international clinical trial protocol registers have been established. In Case 2, central statistical monitoring methods could have assisted in detecting the data falsification. It is interesting to note that in Case 2 the results of the trial remained valid despite the fraudulent data. This was thanks to the design of the trial. The protection against bias afforded by randomisation and blinding which feature in the design of many clinical trials, including this one, protects also against fraudulent activity. Unless the fraud subverts the randomisation process or the blinded assessment of patient outcomes it will have a minimal impact on the final conclusions of the trial analysis.

This does not render the fraud harmless. The public perception of medical research may be greatly damaged by reporting of such cases and can make people unwilling to participate in future research studies. A further spin-off of fraud is that it can lead, as in Case 2, to a substantial increase in audit intensity, consuming resources which could otherwise have been used to focus on the aspects of trials that would improve their

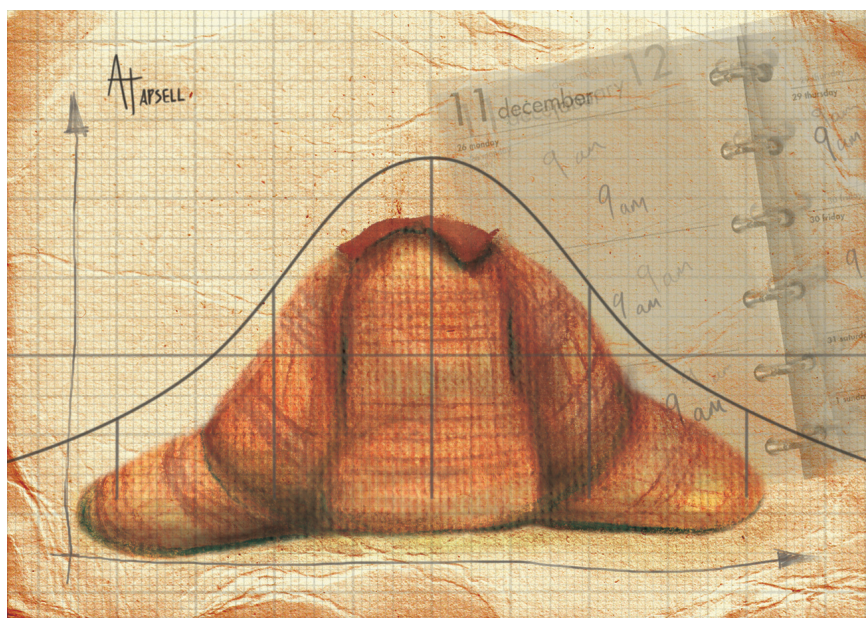


Illustration: Andrew Tapsell, <http://andrewtapsell.blogspot.com>

Box 1. Two widely publicised fraud cases

Case 1

In 2000, Professor Werner Bezwoda¹ of the University of the Witwatersrand, Johannesburg, South Africa, was dismissed after admitting “a serious breach of scientific honesty and integrity” regarding the breast cancer trial results he presented at the May 1999 meeting of the American Society of Clinical Oncology. His trial had reported a beneficial effect of high-dose chemotherapy combined with peripheral blood stem-cell rescue in women with high-risk breast cancer which had spread to several lymph nodes. An on-site review of data from the Bezwoda trial was conducted as part of the planning of a large clinical trial to confirm his potentially important findings. The reviewers were only provided data from three-quarters of the high-dose chemotherapy patients in the trial and none of the control group patients. Only a minority of patients had properly documented evidence of their eligibility for the trial. Most disturbingly, there was no evidence of the consent of patients to participate in the trial and no submission of the trial protocol for approval by the University’s Committee for Research on Human Subjects.

Case 2

By 1994 the National Surgical Adjuvant Breast and Bowel Project (NSABP), chaired by Dr Bernard Fisher, had randomised around 50 000 patients in several dozen clinical trials evaluating strategies for management of early breast cancer, contributing to a substantial body of evidence showing the safety of breast conservation surgery plus radiotherapy relative to the established practice, mastectomy. In 1991 the NSABP had identified that around 100 patients, 0.2% of those randomised, had been included in trials by a Montreal doctor who had altered some of the patients’ details to make them appear eligible. Dates of biopsy and surgery had been changed – the interval between them was one of the eligibility criteria for the trial. Pathology reports and hormone-receptor values had been changed as well. In the suspect cases the patients would have been ineligible if the data had remained unaltered. The NSABP reported this to the National Institutes of Health (NIH) which eventually requested that all of the NSABP trials should be reanalysed without any of the data (about 1500 patients) contributed by the Montreal centre. The NSABP did not prioritise this, presumably due to the lack of scientific justification for omitting data from 1400 eligible and properly documented patients; further, the 100 ineligible patients represented such a small fraction of the overall data that their inclusion would not have biased the clinical findings of the trials.

In 1994, media coverage of the data falsification and subsequent reanalysis delay misleadingly led patients to believe that the clinical results of the NSABP trials were untrustworthy. Two full Congressional Subcommittee hearings summoned the NIH and National Cancer Institute (NCI) directors for an explanation; thereafter, the University of Pittsburgh, at the request of the then NCI director, dismissed Dr Fisher and his senior statistical colleague. A new Clinical Trials Monitoring Branch was created within the NCI to co-ordinate intensive auditing of cancer clinical trials².

ability to answer therapeutic questions reliably. In addition, the increased noise associated with falsified data will reduce the ability of the trial to detect the effects of the treatment being investigated. In a particular type of clinical trial known as an equivalence trial this is critical as it increases the risk of an incorrect conclusion that the treatments are equivalent.

The fraud detective

How can we detect fraud in clinical trials? The conventional approach has been through site visits to medical centres involved in such

trials, checking the source data against what is written down. Any difference between the two may or may not be due to fraud – it might be carelessness, incorrect rounding, or some other malfunction – but shows that further investigation is needed. Checking every item of data against the original source is of course highly time-consuming and costly. Site monitoring therefore often involves random sampling strategies to assess a subset of the entire trial data set.

But, during the trial and after, there are other methods of detection. Perhaps fortunately, it is surprisingly hard to fabricate data

convincingly, as we shall see. All kinds of statistical clues are left which the skilled statistical detective can follow.

The statistical checks can be run during the course of the trial to allow time for any remedial action to be taken, or after it if doubts arise later. Comparing results from different hospital sites or doctors can be informative when fraudulent activity is only being perpetrated in one of several participating hospitals.

Fraud can be plagiarism (not dealt with here), falsification or fabrication. The International Society for Clinical Biostatistics Subcommittee on Fraud has published a detailed overview of these issues in relation to clinical trials³. Areas of clinical trial data collection that are particularly susceptible to fraud include eligibility criteria (for example, the age, sex, and medical history of the patients); measurements such as blood pressure or laboratory data which are requested repeatedly throughout the clinical trial (tedious and time-consuming for both clinician and patient, who may not see the point of them); adverse event reporting (it may be tempting to silently drop from the trial potential evidence of unfortunate side-effects, for instance); assessment of medication compliance based on counting the unused pills; and patient diaries, where one may gain insights into the presence of fraud from the handwriting, colour of ink and texture of pen.

What, then, are the tools that clinical trial statisticians have at their disposal to detect or prevent fraud?

Methods in one dimension

We can obtain insights into the truthfulness of a variable that has been recorded by studying various aspects of its distribution. We can look at the mean and the variability in the measurement for each clinical trialist or medical centre: while individuals inventing data may be able to generate appropriate mean values, they may struggle to create an appropriate level of variability in the faked data. We could set the bar even higher by examining, for instance, the kurtosis (the heaviness of the upper and lower tails of the distribution) too!

Data which include dates of assessment also offer opportunities for verification. It is straightforward to convert dates in a data set to days of the week to check that assessments were performed on plausible days (for example, not on Sundays), or indeed whether the dates correspond to key dates such as public holidays



© iStockphoto.com/Stigur Karlsson

when scheduled assessments would have been unlikely.

Benford's law for first digits⁴ offers a subtle technique for identifying whether a variable has an appropriate distribution. If certain criteria are satisfied, the leading (first significant) digit of a data value turns out to have a somewhat surprising and non-uniform probability distribution. One might expect these digits to be uniformly distributed over the range 1 to 9; in fact, when Benford's law holds, 1 appears as the first digit much more frequently (about 30% of the time), 2 leads about 18% of the time, and there is a steady decrease down to a frequency of less than 5% for a leading digit of 9. This distribution is independent of scale, applying equally to measurements in microlitres or cell counts in thousands. The phenomenon is present in a variety of situations, such as river lengths and the populations of US counties. It does not always apply, but it does remain valid for mixtures of distributions even if inappropriate for some of the variables. In attempting to spot fraudulent activity it may therefore be of greatest use when applied to the combined data from, for example, all of the laboratory measurements in a clinical trial. Used in this way it would be most useful as a tool for identifying fraud on a large scale from a single triallist or medical centre in a multi-centred trial.

In addition to the leading digit strategy, we can explore the terminal digit preference for a variable. This of course, unlike the leading digit, should be evenly spread across the possible values. In addition to this type of formal statistical testing, we can apply com-

Benford's law says that first digits should be weighted towards 1, 2 and 3; last digits should be random

mon sense checks: if the equipment used to assess lung performance in asthmatic patients reports whole-number results then if the data set records digits after the decimal point those digits must have been invented.

The multi-dimensional perspective

Clinical trials gather data on several measurements on each patient: rather than considering each in isolation, adopting a multi-dimensional approach allows us to exploit the subtle interrelationships that exist between variables. If it is hard convincingly to fabricate a single variable, it is much harder to make a set of several interrelated measurements look genuine.

In conventional statistical analysis there is often a focus on detecting outlying observations – those which may be having an undue influence on the overall conclusions. In the detection of fabricated data, as well as outliers we are also particularly concerned with identifying “inliers”: patients for whom the set of measurements made lie too close to the overall mean. The rationale is that those inventing data are more likely to place their fabricated data values close to the mean in order to make them less obvious and avoid detection.

The step-by-step guide to inlier detection is given in Box 2.

Cluster analysis⁵ may be used to detect suspected duplicate samples, which would occur, for example, if the researcher had taken a single blood sample and split it before sending it to the laboratory to obtain a range of biochemistry measurements. Here, the results from the parts of the sample would be too similar to have occurred by chance.

Tracking measurements over time

In many clinical trials patients are followed up over repeated visits. This gives further opportunities for data checking. Patients are human beings, and human beings are various and variable. How many of us, if called to a hospital appointment every Monday, would be able to

Box 2. Detecting inliers

Detecting data that are fabricated, using inliers, can be done using the following procedure:

- 1 Standardise the observations on each variable, by subtracting the mean and dividing by the standard deviation:
 $Z = (x - \mu) / \sigma$.
- 2 For each patient, add up the Z^2 values across the p variables (an enhanced approach would also take account of correlations between variables to calculate what is known as the Mahalanobis distance).
- 3 This sum should approximately follow the χ^2 distribution with p degrees of freedom.
- 4 The distribution of summed Z^2 values may be plotted on a log scale (which enables easy identification of unexpectedly small values), annotated by clinician or study site.

The above approach readily allows inliers – patients for whom the summed Z^2 values are smaller than would be expected by chance – to be highlighted, as shown in Figure 1.

keep the schedule unbroken for six months or more? There are holidays, there are birthdays, there are work or family commitments. Often the schedule assigned to each patient takes this into account and permits some deviation, for example ± 1 day for weekly visits or ± 1 week for six-monthly visits to allow for holidays or other valid reasons for non-attendance. For patients seen at a given medical centre or by a particular doctor, excessive instances of perfect attendance on the scheduled day could be a hallmark of falsified data (Figure 2). The distribution of the intervals between visits for a patient may be a further indicator of fake data if the time gap between visits is too consistent.

The values of the repeated measurements themselves may also provide clues. The kurtosis of the distribution of differences in repeated measures may flag up fabricated data where the clinical researcher has copied the results from one visit to the next while changing the value by some arbitrary amount to avoid detection. Changes in the variance of a measurement over time may highlight the start (or end!) of a period of data fabrication.

The integrity of the randomisation process is key to the validity of a clinical trial. To this end, some simple checks should be used to verify that time trends are absent from the proportion of patients allocated to each treatment or placebo. You would not, for

example, want to see that the patients allocated active treatment and placebo at the start had a similar random spread of blood pressures, while those allocated to active treatment later

Patients do not always keep appointments. A perfect attendance record may indicate fraud

on all had higher blood pressures. This would suggest that a triallist had gained knowledge of the identity of the treatments in a blinded trial and had subsequently used this when including patients in the trial. Randomisation dates can be subjected to the same common-sense validity checks as for dates in general in the one-dimensional methods section above.

Prevention

Although it may be impossible to eradicate fraud completely, are there any steps we might take to minimise its occurrence? It will pay to focus on the data items that are most easily affected by fraud, such as eligibility criteria, repeated measurements, adverse events, compliance data and patient diaries.

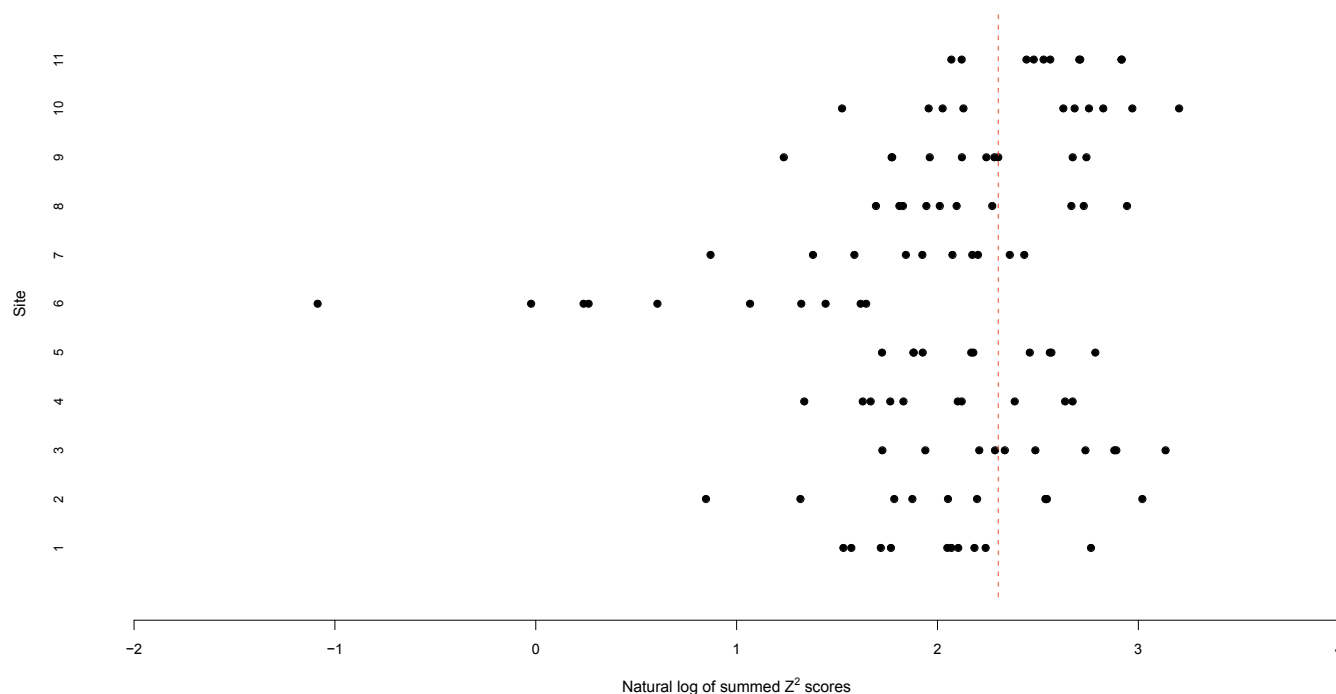


Figure 1. Fictitious data illustrating the investigation of possible inliers in a trial where patients have been studied at 11 medical centres (sites) and 10 variables have been recorded on each patient. The logged summed Z^2 values for patients from site 6 are rather smaller than their expected value (red dotted line), indicating the presence of inliers

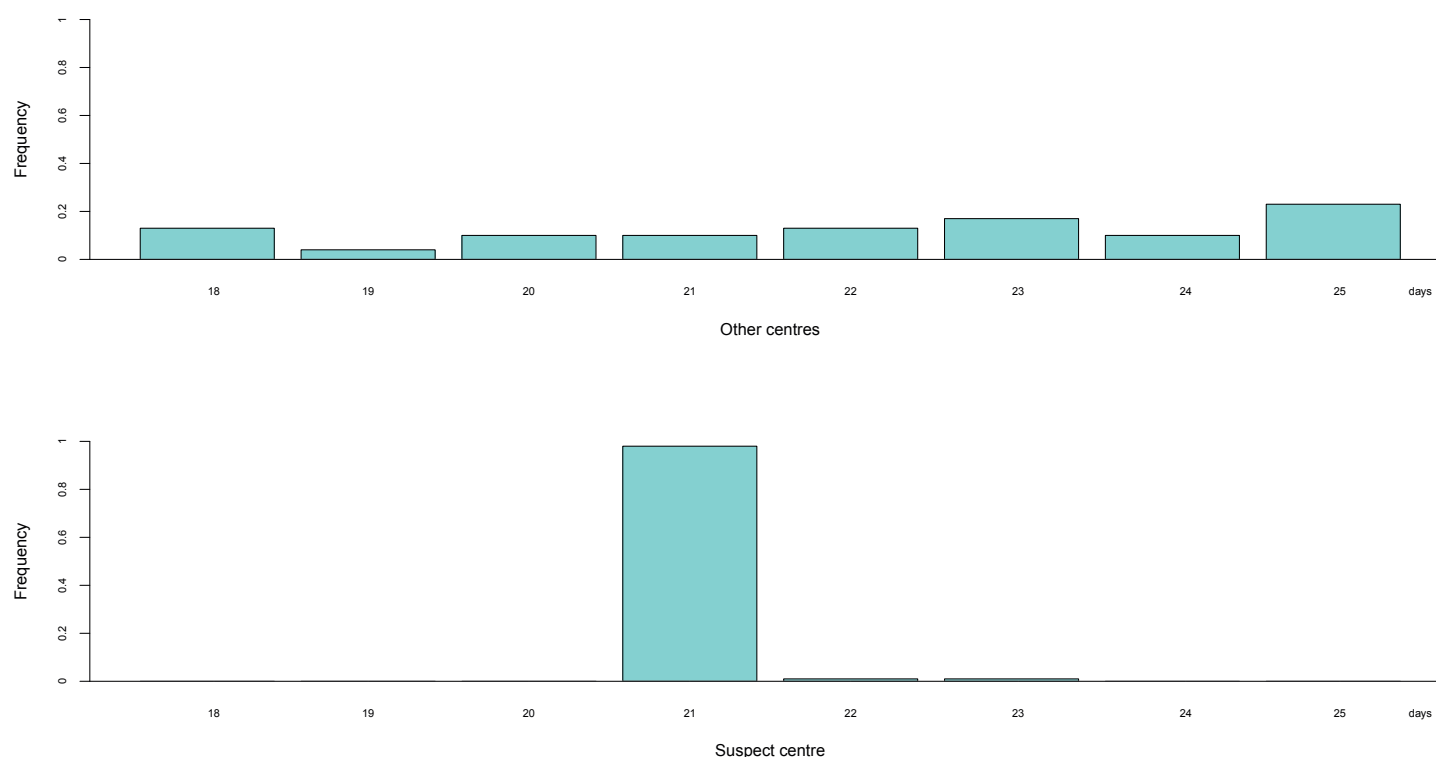


Figure 2. Near-perfect attendance on scheduled visit date raises concern regarding one study site. Based on Table IV data from Buyse *et al.*³

What might motivate fraudulent activity? Data may be fabricated to modify the conclusions to the desired results; for financial benefit in cases where doctors are paid money for each patient they enrol in a clinical trial; for the gain in prestige that may come from being involved in an international clinical trial; or due to laziness, for example where repeated assessments of blood pressure are required within a clinical trial protocol. Not all fraud stems from moral turpitude: it may result from compassion on the part of doctors, whereby they include in a trial a patient whom they believe would gain benefit, despite the patient being ineligible: perhaps this was the reason for the falsification by the Montreal doctor in Case 2 (Box 1).

Two potentially successful preventive strategies are simplifying the eligibility criteria for the trial and reducing the numbers of variables being recorded. These measures are both feasible and mostly do not impair the validity of trial findings. In large clinical trials, embedding assessments as far as possible within routine clinical practice⁶ is one way to encourage adherence to the protocol by triallists.

If every missing data item in a trial were queried by the trial organisers, regardless of its importance, then this would clearly put temptation in the way of the triallist to fabricate

data in order to avoid having to respond to long lists of queries. Selection of items for querying and monitoring should therefore be based on risk assessment to identify the critical variables; central statistical monitoring is likely to play a central role in the strategy devised for a given clinical trial⁷.

Conclusions

The protective features of randomisation and blinding that are present in many clinical trials do limit the possible impact of fraudulent activity on the final trial conclusions. Applying a selection of the statistical methods we have outlined, at regular intervals during the course of a clinical trial, would also enable prompt detection and resolution of any fraudulent activity. This would remove one source of excess noise in the data. More vitally, it would preserve the public perception of research integrity.

References

1. Horton, R. (2000) Commentary: After Bezwoda. *Lancet*, **353**, 942–943.
2. Peto, R., Collins, R., Sackett, D. *et al.* (1997) The trials of Dr Bernard Fisher: a European perspective on an American episode. *Controlled Clinical Trials*, **18**, 1–13.

3. Buyse M, George, S.L., Evans, S. *et al.* for the ISCB Subcommittee on Fraud (1999) The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, **18**, 3435–3451.

4. Benford, F. (1938) The law of anomalous numbers. *Proceedings of the American Philosophical Society*, **78**, 551–572.

5. Wu, X. and Carlsson, M. (2011) Detecting data fabrication in clinical trials from cluster analysis perspective. *Pharmaceutical Statistics*, **10**, 257–264.

6. Warlow, C.P. (2002) Advanced issues in the design and conduct of randomized clinical trials: the bigger the better? *Statistics in Medicine*, **21**, 2797–2805.

7. Baigent, C., Harrell, F.E., Buyse, M., Emberson, J.R. and Altman, D.G. (2008) Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials*, **5**, 49–55.

Gordon Murray is Professor of Medical Statistics at the University of Edinburgh and Director of the MRC Edinburgh Hub for Trials Methodology Research (HTMR). Christopher Weir is Associate Director (Statistics) of the MRC Edinburgh HTMR and the Edinburgh Health Services Research Unit. The authors were supported in part in this work by the UK Medical Research Council Network of Hubs for Trials Methodology Research. The views expressed are those of the authors and not necessarily of organisations with which the authors are affiliated.