Al-Marzouki, S; Evans, S; Marshall, T; Roberts, I (2005) Are these data real? Statistical methods for the detection of data fabrication in clinical trials. BMJ (Clinical research ed), 331 (7511). pp. 267-70. ISSN 0959-8138 DOI: https://doi.org/10.1136/bmj.331.7511.267

Downloaded from: http://researchonline.lshtm.ac.uk/13500/

DOI: 10.1136/bmj.331.7511.267

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: Creative Commons Attribution Non-commercial http://creativecommons.org/licenses/by-nc/3.0/
Are these data real? Statistical methods for the detection of data fabrication in clinical trials

Sanaa Al-Marzouki, Stephen Evans, Tom Marshall, Ian Roberts

Abstract
Objectives To test the application of statistical methods to detect data fabrication in a clinical trial.
Setting Data from two clinical trials: a trial of a dietary intervention for cardiovascular disease and a trial of a drug intervention for the same problem.
Outcome measures Baseline comparisons of means and variances of cardiovascular risk factors; digit preference overall and its pattern by group.
Results In the dietary intervention trial, variances for 16 of the 22 variables available at baseline were significantly different, and 10 significant differences were seen in means for these variables. Some of these P values were extraordinarily small. Distributions of the final recorded digit were significantly different between the intervention and the control group at baseline for 14/22 variables in the dietary trial. In the drug trial, only five variables were available, and no significant differences between the groups for baseline values in means or variances or digit preference were seen.
Conclusions Several statistical features of the data from the dietary trial are so strongly suggestive of data fabrication that no other explanation is likely.

Introduction
Most statistical analyses of clinical trials are undertaken on the presumption that the data are genuine. Large accidental errors can be detected during data analysis,1 2 but if people are trying to “make up” data they are likely to do it in such a way that it is not immediately obvious, avoiding any large discrepancies. Nevertheless, fraudulent data have particular statistical features that are not evident in data containing accidental errors, and several analytical methods have been developed to detect fraud in clinical trials.3 4 The BMJ has taken a general interest in this field and has published a book on fraud and misconduct, now in its third edition, which has a chapter on statistical methods of detection of fraud.5

In this paper we use statistical techniques to examine data from two randomised controlled trials. In one trial, the possibility of scientific misconduct had been raised by BMJ referees, based on inconsistencies in calculated P values compared with the means, standard deviations, and sample sizes presented (see p 281). For comparison, we used the same methods to analyse a second trial for which there were no such concerns. We were not involved in either trial.

Methods
The trial about which doubts were raised (the diet trial) was a single blind, randomised controlled trial of the effects of a fruit and vegetable enriched diet in 831 patients with coronary heart disease, including patients with angina pectoris, myocardial infarction, or surro-
Table 1 Baseline variables in the two trials under comparison

<table>
<thead>
<tr>
<th>Variable</th>
<th>Diet</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (kg):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>65.74</td>
<td>65.59</td>
</tr>
<tr>
<td>Median</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Mode</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>SD</td>
<td>7.89</td>
<td>7.64</td>
</tr>
<tr>
<td>Min</td>
<td>40</td>
<td>59</td>
</tr>
<tr>
<td>Max</td>
<td>87</td>
<td>85</td>
</tr>
<tr>
<td>Height (cm):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>165.1</td>
<td>165.28</td>
</tr>
<tr>
<td>Median</td>
<td>162</td>
<td>160</td>
</tr>
<tr>
<td>Mode</td>
<td>165</td>
<td>165</td>
</tr>
<tr>
<td>SD</td>
<td>3.93</td>
<td>3.93</td>
</tr>
<tr>
<td>Min</td>
<td>142</td>
<td>140</td>
</tr>
<tr>
<td>Max</td>
<td>179</td>
<td>178</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>134.2</td>
<td>131.9</td>
</tr>
<tr>
<td>Median</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>Mode</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>SD</td>
<td>16.3</td>
<td>16.9</td>
</tr>
<tr>
<td>Min</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Max</td>
<td>200</td>
<td>195</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>86.5</td>
<td>86.7</td>
</tr>
<tr>
<td>Median</td>
<td>86</td>
<td>85</td>
</tr>
<tr>
<td>Mode</td>
<td>80</td>
<td>85</td>
</tr>
<tr>
<td>SD</td>
<td>9.7</td>
<td>9.7</td>
</tr>
<tr>
<td>Min</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Max</td>
<td>112</td>
<td>110</td>
</tr>
<tr>
<td>Cholesterol (mmol/l):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.46</td>
<td>5.43</td>
</tr>
<tr>
<td>Median</td>
<td>5.48</td>
<td>5.48</td>
</tr>
<tr>
<td>Mode</td>
<td>5.43</td>
<td>5.43</td>
</tr>
<tr>
<td>SD</td>
<td>0.52</td>
<td>0.296</td>
</tr>
<tr>
<td>Min</td>
<td>4.53</td>
<td>2.95</td>
</tr>
<tr>
<td>Max</td>
<td>6.52</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Results

Table 1 shows descriptive summaries of variables common to both trials for both groups in each trial. The drug trial values show what might be expected in a randomised trial, but the diet trial shows notable differences in standard deviations for height and cholesterol measurements.

Discussion

The data from the diet trial have various anomalous statistical features that are not present in the data from the drug trial. These features are differences in means, and, even more noticeable, in variances at baseline and in differences in pattern of digit preference between randomised groups.

Magnitude of P values

These differences in the means and variances between baseline variables in the diet trial indicate that the two groups simply cannot have been formed as a result of random allocation as the authors claim. The magnitude of the P values derived from t tests of these differences for several variables is not compatible with a chance effect. One or two variables might show a small effect, but several of these P values are extreme.
Similarly, the significant difference in the pattern of digit preference between the randomised groups provides additional evidence that this is not a truly randomised trial.

**Randomisation process**

If this is not a randomised trial then how did these data arise? One possibility is that the data themselves are genuine but that the randomisation process has been subverted. This might explain, for example, some of the differences between the means of the variables at baseline. Had there been subversion of the randomisation process, in order for example to create differences between the groups at baseline, then smaller differences would have occurred and would also have been more consistent between the variables that are medically related—such as the different meas-

### Table 2 Baseline comparison of the two intervention groups, diet trial and drug trial

<table>
<thead>
<tr>
<th></th>
<th>Diet trial</th>
<th>Drug trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F Significance</td>
<td>t test for equality of means</td>
</tr>
<tr>
<td>Height</td>
<td>71.15</td>
<td>1.4x10^{-11}</td>
</tr>
<tr>
<td>Weight</td>
<td>0.204</td>
<td>0.652</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>4.81</td>
<td>0.029</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>28.77</td>
<td>1x10^{-1}</td>
</tr>
<tr>
<td>Fasting blood glucose</td>
<td>8.21</td>
<td>0.004</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>0.043</td>
<td>0.835</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>21.98</td>
<td>3x10^{-4}</td>
</tr>
<tr>
<td>Energy</td>
<td>0.98</td>
<td>0.322</td>
</tr>
<tr>
<td>Total carbohydrate</td>
<td>1.97</td>
<td>0.161</td>
</tr>
<tr>
<td>Complex carbohydrate</td>
<td>12.86</td>
<td>0.0004</td>
</tr>
<tr>
<td>Protein</td>
<td>15.18</td>
<td>0.0002</td>
</tr>
<tr>
<td>Fat</td>
<td>20.5</td>
<td>7x10^{-3}</td>
</tr>
<tr>
<td>Saturated</td>
<td>15.2</td>
<td>0.0001</td>
</tr>
<tr>
<td>Fibre</td>
<td>94.23</td>
<td>4x10^{-31}</td>
</tr>
<tr>
<td>Soluble fibre</td>
<td>10.13</td>
<td>0.002</td>
</tr>
<tr>
<td>Caffeine</td>
<td>2.41</td>
<td>0.121</td>
</tr>
<tr>
<td>Salt</td>
<td>39.72</td>
<td>5x10^{-6}</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>0.007</td>
<td>0.951</td>
</tr>
<tr>
<td>Carotene</td>
<td>51.06</td>
<td>2x10^{-18}</td>
</tr>
<tr>
<td>Vitamin E</td>
<td>25.7</td>
<td>5x10^{-4}</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>51.42</td>
<td>2x10^{-18}</td>
</tr>
</tbody>
</table>

The χ² value has 9 degrees of freedom.

### Table 3 χ² value (with P value) for the final digit at baseline, diet trial and drug trial

<table>
<thead>
<tr>
<th></th>
<th>Diet trial</th>
<th>Control</th>
<th>Drug trial</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>—</td>
<td>—</td>
<td>239 (1.8x10^{-4})</td>
<td>251 (7.2x10^{-14})</td>
</tr>
<tr>
<td>Weight</td>
<td>128 (4x10^{-3})</td>
<td>23 (0.00655)</td>
<td>7.3 (0.60)</td>
<td>6.5 (0.69)</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>1786 (U)</td>
<td>1470 (U)</td>
<td>7.6 (0.58)</td>
<td>8.1 (0.43)</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>554 (2x10^{-10})</td>
<td>430 (6x10^{-10})</td>
<td>16.23 (0.062)</td>
<td>5.76 (0.76)</td>
</tr>
<tr>
<td>Fasting blood glucose</td>
<td>478 (4x10^{-9})</td>
<td>538 (5x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>1053 (6x10^{-11})</td>
<td>1322 (U)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>642 (2x10^{-12})</td>
<td>963 (2x10^{-12})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Energy</td>
<td>2101 (U)</td>
<td>2630 (U)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total carbohydrates</td>
<td>2671 (1x10^{-9})</td>
<td>927 (3x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Complex carbohydrates</td>
<td>231 (1x10^{-9})</td>
<td>939 (3x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Protein</td>
<td>54 (2x10^{-4})</td>
<td>251 (5x10^{-4})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fat</td>
<td>229 (2x10^{-4})</td>
<td>437 (2x10^{-4})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Saturated</td>
<td>123 (4x10^{-10})</td>
<td>98 (4x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fibre</td>
<td>263 (2x10^{-10})</td>
<td>1127 (8x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Soluble fibre</td>
<td>213 (1x10^{-10})</td>
<td>1086 (8x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Caffeine</td>
<td>813 (3x10^{-10})</td>
<td>684 (1x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Salt</td>
<td>288 (9x10^{-4})</td>
<td>301 (2x10^{-3})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>364 (5x10^{-4})</td>
<td>411 (6x10^{-4})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Carotene</td>
<td>1470 (U)</td>
<td>1156 (5x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Vitamin E</td>
<td>118 (3x10^{-10})</td>
<td>101 (8x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>705 (8x10^{-10})</td>
<td>799 (3x10^{-10})</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

The χ² value has 9 degrees of freedom.

* U means that the P value is too small for calculation.
The degrees of freedom are less than 9 when one or more digits do not appear.

| Table 4 | $\chi^2$ value (with P value) for the final digit at the baseline in the diet and drug trials between the two randomised groups |
|-----------------|-----------------|-----------------|-----------------|
|                | Diet trial      |                 | Drug trial      |                 |
|                 | $\chi^2$test (P value) | df | $\chi^2$test (P value) | df |
| Height          | —               | —               | 5 (0.83)        | 9               |
| Weight          | 36 (3x10$^{-1}$) | 9               | 10 (0.31)       | 9               |
| Systolic blood pressure | 26 (0.00018) | 9               | 7 (0.69)        | 9               |
| Diastolic blood pressure | 16 (0.046) | 9               | 10 (0.38)       | 9               |
| Cholesterol     | 13 (0.182)      | 9               | 7 (0.69)        | 9               |
| Fasting blood glucose | 12 (0.2) | 9               | —               | —               |
| Total cholesterol | 46 (5x10$^{-1}$) | 9 | —               | —               |
| Triglycerides    | 48 (3x10$^{-1}$) | 9 | —               | —               |
| Energy           | 16 (0.064)      | 9 | —               | —               |
| Total carbohydrate | 154 (2x10$^{-1}$) | 9 | —               | —               |
| Complex carbohydrate | 135 (1.4x10$^{-1}$) | 9 | —               | —               |
| Protein          | 43 (2x10$^{-1}$) | 9 | —               | —               |
| Fat              | 40 (6.4x10$^{-1}$) | 9 | —               | —               |
| Saturated        | 15 (0.08)       | 9 | —               | —               |
| Fibre            | 157 (8x10$^{-2}$) | 8 | —               | —               |
| Soluble fibre    | 175 (6.5x10$^{-2}$) | 9 | —               | —               |
| Caffeine         | 15 (0.059)      | 9 | —               | —               |
| Salt             | 28.5 (0.001)    | 9 | —               | —               |
| Vitamin C        | 18 (0.03)       | 9 | —               | —               |
| Carotene         | 10 (0.266)      | 8 | —               | —               |
| Vitamin E        | 20 (0.017)      | 9 | —               | —               |
| Vitamin A        | 9.5 (0.4)       | 9 | —               | —               |

The degrees of freedom are less than 9 when one or more digits do not appear.

Diabetes preference

Diabetes preference in itself is not evidence of misconduct. It is conceivable that the different patterns of diabetes preference between the two randomised groups may have arisen had one person recorded data for the treatment group and another recorded data for the control group. However, it is claimed that the trial was single blind, meaning that those recording data should not know to which group patients had been allocated. We would not expect differences therefore in diabetes preference between the randomised groups. But perhaps the trial was not single blind as described, and those recording the data were separated into groups according to whether they were dealing with patients allocated to either treatment or control. This could lead to differences in diabetes preference between randomised groups for variables where a human element of judgment was required. This would still not explain the differences in means and variances between the two groups since the effect of diabetes preference on the means and variances would only be slight. The combination of the differences in means, variances, and diabetes preference between the randomised groups is strong evidence that data fabrication took place in the diet trial.

Conclusion

We conclude that the data from the diet trial were either fabricated or falsified and that the strength of the evidence is such that appropriate steps should be taken to deal with this matter.

We thank Tom Meade who, on behalf of the Medical Research Council, provided the data for the drug trial and Richard Smith for his encouragement to examine further the data from the diet trial. The BMJ provided the data from the diet trial, which were supplied by the original author for further investigation of these data.

Contributors: SE and SAM had the ideas for the analysis, and SAM, SE, TM, and IR all contributed to the planning, conduct, and writing of the paper. SAM planned and carried out the statistical analyses. SAM and SE are jointly responsible for the overall content as guarantors. There are no other contributors. Competing interests: None declared.

Funding: None.

What is already known on this topic

Data fabrication is a rare form of scientific misconduct in clinical trials, but when it does occur it has serious consequences.

Most papers are published without their data being independently verified, and there have been calls for data to be made available for scrutiny.

Statistical methods for the detection of misconduct have been described, but few examples of their application have been published.

It has been stated that statistical methods alone cannot prove data fabrication.

What this study adds

Statistical methods can be applied to detect large scale fabrication of data in a randomised trial where data are available.

Certain patterns of data are incompatible with randomisation, especially when a trial is "blind".

This paper shows the fabrication or falsification of data in a particular trial.