

Sloppy stats shame science

Jun 3rd 2004

From *The Economist* print edition

What is published in scientific journals may not be as true as it should be

SCIENTIFIC and medical journals, with their august panels of peer reviewers and fact checkers, are not the sort of places many mistakes are to be expected. Yet Emili García-Berthou and Carles Alcaraz, two researchers at the University of Girona in Spain, have found that 38% of a sample of papers in *Nature*, and a quarter of those sampled in the *British Medical Journal* (BMJ)—two of the world's most respected journals—contained one or more statistical errors. Not all of these errors led to erroneous conclusions, but the authors of the study, which has just been published in *BMC Medical Research Methodology*, another journal, reckon that 4% of the errors may have caused non-significant findings to be misrepresented as being significant.

Dr García-Berthou and Dr Alcaraz investigated 32 papers from editions of *Nature* published in 2001 and 12 from the BMJ in the same year. They examined the numbers within each, to see whether the data presented actually led to the statistical conclusion the authors drew, and whether there was anything fishy about the numbers themselves. Appropriately, they used a statistical technique to do their checking. If a set of data are “unedited”, the last digits in the numbers recorded will tend to have the values 0-9 at random, since these digits represent small values, and are thus the ones that are hardest to measure. If those numbers are rounded carelessly, however, 4s and 9s (which tend to get rounded up to the nearest half or whole number) will be rarer than they should be. The two researchers duly discovered that 4s and 9s were, indeed, rarer than chance would predict in many of the papers under scrutiny.

False data, false results. Though it was difficult to show whether, in any given case, this falsity led to a result being proclaimed statistically significant when it was not, it was possible to estimate how much error there was likely to be. In one case, however, there was no doubt. A number supposed to be statistically significant was explicitly mis-stated, and a false inference drawn in the paper's conclusion.

Of course, mistakes will creep through from time to time in the best-run organizations, and there is no suggestion that any of the errors observed was a deliberate fraud. But there do seem to have been rather a lot of them. However, as Kamran Abbasi, deputy editor of the BMJ, laments, although the world at large looks at scientific peer-review—the system journals use to keep their authors accurate and honest—as a sacred process, it is in fact imperfect. “We certainly do not spend our time recalculating all these numbers, and our whole review process would likely grind to a halt if we tried to do so.”

Maxine Clarke, publishing executive editor of *Nature*, says her journal will be examining the papers cited by Dr García-Berthou and Dr Alcaraz before deciding what action, if any, needs to be taken. “At first sight, some awareness raising about statistical accuracy among manuscript editors, peer-reviewers and proofreaders seems necessary, but we have changed our work flows considerably since the period studied,” says Ms Clarke.

One cure might be for researchers to publish raw data as well as statistical analysis and conclusions. That way, anyone who really cares can check the sums. For some years, *Nature* has offered supplementary information online to accompany its papers. This information is peer-reviewed, but Ms Clarke believes it is too specialized for people outside the field to find interesting. “We do not explicitly ask authors, as routine, for the raw data underlying their reported statistical results,” she says. “This suggestion is now on the agenda for our next editorial meeting on editorial practices and criteria.”

The real answer, however, surely lies with the researchers themselves. Far too many scientists have only a shaky grasp of the statistical techniques they are using. They employ them as an amateur chef employs a cookbook, believing the recipes will work without understanding why. A more cordon bleu attitude to the maths involved might lead to fewer statistical soufflés failing to rise.