

Statistically significant

Last year, *Nature* and *Nature Medicine* were publicly criticized for what could be described as the sloppiness of the statistical analysis in some of our published articles. These criticisms prompted us to take a close look at the statistical methodology used in our papers, as we tried to determine the true extent of the problem and whether we needed to devise a system to prevent it from recurring. The results of this soul-searching exercise turned out to be very instructive.

Our concerns began last May with the publication of a paper in *BMC Medical Research Methodology* (4, 13 (2004)). In their article, Emili García-Berthou and Carles Alcaraz, from the University of Girona, Spain, checked the accuracy of the statistical results reported in the 181 research papers that *Nature* published during 2001 and found that 38% of the articles contained at least one statistical error. The authors concluded that quality control of scientific papers needs to be more carefully monitored and suggested that a way to minimize these errors would be for published authors to make their raw data freely available on the Internet.

The findings of this study captured the attention of the media, leading to a series of reports in the international press. One of them, written by Robert Matthews for *The Financial Times*, went one step further than reporting the results, and included an original analysis of the statistical methodology of *Nature Medicine* papers published in 2000. Matthews found that 31% of our articles showed evidence that their authors misunderstood the meaning of *P* values, leading to, for example, reports of *P* with ludicrous precision (e.g. $P = 0.002387$).

How serious is this problem? To answer this question, we decided to commission an independent 'statistical audit' of *Nature Medicine* papers from two Columbia University experts. Specifically, we asked them to review the statistical methods of a subset of our material, the 21 articles involving human subjects that we published during 2003.

Using a checklist of commonly accepted statistical reporting criteria, the two statisticians evaluated the papers and concluded that their authors had a wide range of statistical expertise. At one end of the spectrum, some papers had almost no quantitative analysis. At the other end, some included rather sophisticated statistical and mathematical methodology. But most of the articles fell in the middle, containing a few statistical tests to support the authors' interpretation of the data.

These tests were often incompletely described, making it difficult to assess their appropriateness to analyze the sample under scrutiny.

Some of the omissions that the analysis disclosed were frankly surprising, owing to their apparent simplicity. Authors often failed to state the sample size, and occasionally introduced rounding and truncation errors. They frequently reported *P* values while failing to mention the statistical tests they used to obtain them. In some cases, the statistical measures were not labeled, making it impossible to establish whether they represented standard deviation or standard error. And in some cases in which the standard deviation or error was identified as such, the sample size was too small to warrant its calculation.

As is evident, these problems are largely the result of inadequate provision of detail and do not even begin to explore *bona fide* statistical errors (which were also common): the use of one-tailed instead of two-tailed tests or the lack of adjustment of the level of statistical significance in the case of multiple pairwise comparisons. In short, what we learned from this audit was that the statistical sophistication of most of our authors, referees and editors is rather elementary, and that the criticisms that we received are legitimate, requiring us to take prompt action.

So, as a result of our independent audit, we have decided to take steps towards improving the quality of the statistical reporting in *Nature Medicine*. Reflecting on the types of errors that the audit disclosed, we concluded that for most papers, the problems encountered are quite basic and would not warrant a full review by a statistician, as is common practice in clinical journals. Instead, we believe that the common errors that we found can be remedied by enforcing clear guidelines about descriptions of quantitative data and statistics. We are in the process of finalizing these guidelines, which will appear in our Guide to Authors within the next few weeks.

The guidelines, which will ultimately be adopted by *Nature* and all the Nature Research journals, will require authors to include a subsection on statistics in the Methods section of their papers, and will include a discussion of at least three broad topics: statistical testing, descriptive statistics and common statistical errors. We are confident that the guidelines will assist not only the authors in preparing their manuscripts, but also the editors and referees in evaluating the validity of the data.