Reviewer's report

Title: Error in statistical tests of error in statistical tests

Version: 1 Date: 22 February 2006

Reviewer: Damien Jolley

Reviewer's report:

General
This paper points out a statistical flaw in an earlier BMC Methods article which reported apparently non-random distributions of P-values in publications from Nature and BMJ. As a statistician myself, I was pleased for the opportunity to review this manuscript.

The publicity following the original article was wide-spread and very negative ("Sloppy Stats Shame Science", said The Economist). If, as the present paper suggests, the statistics in the accusing article were flawed, the implications for scientific publishing are intriguing to consider...

So, can we validate the claims made in this recent submission? I have to confess to a degree of disbelief on first reading the ms to find that an apparently highly significant test result (Z=2.71, P < 0.0005) could be so radically wrong and that the valid P-value for the test of goodness-of-fit is (gasp!) Z=1.1, P=0.09 !!

The apparent seat of the disparity in results would appear to be the nature of the underlying theoretical "uniform" distribution function used by the two sets of authors. The current paper claims that the original was erroneously based on a continuous uniform distribution, and that the correct distribution should be the discrete uniform distribution.

To verify the claim, I needed to reproduce both results using validated software (or write the code myself).

My usual tool of choice (Stata) produced a result identical to the original value, with no discussion in the Manual about the distinction between discrete or continuous distributions.

Happily, I turned next to an excellent utilitarian software package StatXact (Version 6 with Cytel Studio, Cytel Statistical Software, Cambridge MA, 2004). This program offers BOTH "uniform continuous" AND "uniform discrete" options for the Kolmogorov-Smirnov goodness-of-fit test. Using this program, I was able to reproduce the original Z=2.71 (P < 0.0005) using "Uniform continuous" and then Z=1.05 (P=0.106, using Monte Carlo methods).

There is no doubt in my mind that the "Uniform discrete" distribution is the correct choice in this instance. StatXact has confirmed the claim made in this paper, that the original claims were incorrect, and that indeed the collection of last digits from P-values reported in Nature could very reasonably have arisen from a discrete uniform distribution.

Other comments contained in the manuscript are sensible and apposite.

Finally, I add my voice of support to the opinion expressed in the paper's final paragraph. I fail to appreciate the point of worrying about the last digit in reported P-values (or even P-values themselves, for that matter!), and particularly so when the critics' methods are wrong.
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

---

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

---

Discretionary Revisions (which the author can choose to ignore)

---

What next?: Accept without revision

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: Yes

Declaration of competing interests:

I declare that I have no competing interests.