theguardian

The statistical error that just keeps on coming

The same statistical errors – namely, ignoring the "difference in differences" – are appearing throughout the most prestigious journals in neuroscience

Ben Goldacre

guardian.co.uk, Friday 9 September 2011 20.59 BST

In a hypothetical trial on two kinds of mice, several tests are needed to prove statistically significant differences. Photograph: Sam Yeh/AFP/Getty Images

We all like to laugh at quacks when they misuse basic statistics. But what if academics, en masse, deploy errors that are equally foolish? This week Sander Nieuwenhuis and colleagues publish a mighty torpedo in the journal Nature Neuroscience.

They've identified one direct, stark statistical error so widespread it appears in about half of all the published papers surveyed from the academic neuroscience research literature.

To understand the scale of this problem, first we have to understand the error. This is difficult, and it will take 400 words of pain. At the end, you will understand an important aspect of statistics better than half the professional university academics currently publishing in the field of neuroscience.

Let's say you're working on nerve cells, measuring their firing frequency. When you drop a chemical on them, they seem to fire more slowly. You've got some normal mice and some mutant mice. You want to see if their cells are differently affected by the chemical. So you measure the firing rate before and after applying the chemical, first in the mutant mice, then in the normal mice.

When you drop the chemical on the mutant mice nerve cells, their firing rate drops, by 30%, say. With the number of mice you have this difference is statistically significant, and so unlikely to be due to chance. That's a useful finding, which you can maybe publish. When you drop the chemical on the normal mice nerve cells, there is a bit of a drop, but not as much – let's say 15%, which doesn't reach statistical significance.

But here's the catch. You can say there is a statistically significant effect for your chemical reducing the firing rate in the mutant cells. And you can say there is no such statistically significant effect in the normal cells. But you can't say mutant and normal cells respond to the chemical differently: to say that, you would have to do a third statistical test, specifically comparing the "difference in differences", the difference between the chemical-induced change in firing rate for the normal cells against the chemical-induced change in the mutant cells.

Now, looking at the figures I've given you here (for our made up experiment) it's very likely that this "difference in differences" would not be statistically significant, because the responses to the chemical only differ from each other by 15%, and we saw earlier that a drop of 15% on its own wasn't enough to achieve statistical significance.

But in just this situation, academics in neuroscience papers routinely claim to have found a difference in response, in every field imaginable, with all kinds of stimuli and interventions: comparing younger versus older participants; in patients against normal volunteers; between different brain areas; and so on.

How often? Nieuwenhuis looked at 513 papers published in five prestigious neuroscience journals over two years. In half the 157 studies where this error could have been made, it was. They broadened their search to 120 cellular and molecular articles in Nature Neuroscience, during 2009 and 2010: they found 25 studies committing this fallacy, and not one single paper analysed differences in effect sizes correctly.

These errors are appearing throughout the most prestigious journals for the field of neuroscience. How can we explain that? Analysing data correctly, to identify a "difference in differences", is a little tricksy, so thinking generously, we might suggest that researchers worry it's too longwinded for a paper, or too difficult for readers. Alternatively, less generously, we might decide it's too tricky for the researchers themselves.

- But the darkest thought of all is this: analysing a "difference in differences" properly is much less likely to give you a statistically significant result, and so it's much less likely to produce the kind of positive finding you need to look good on your CV, get claps at conferences, and feel good in your belly. Seriously: I hope this is all just incompetence.
  • This article was amended on 13 September 2011 to make clear that the Nieuwenhuis study looked specifically at neuroscience papers, not psychology research.