Faith in science is undermined by peer-review failings

By Judith Sloan, The Australian, 21 October 2018

Science has been in the news lately. As part of the release of the latest UN Intergovernmental Panel on Climate Change report, the boast was made that the contents were based on the work of 91 of the top scientists and more than 6000 scientific references.

This carries on the tradition outlined by the chairman of the IPCC from 2002 to 2015, Rajendra Pachauri: "We carry out an assessment of climate change based on peer-reviewed literature, so everything that we look at and take into account in our assessments has to carry the credibility of peer-reviewed publications, we don't settle for anything less than that."

The trouble for the IPCC — and for many other outlets that carry scientific findings — is that a crisis in science has been brewing for some time. Known as the replication or reproducibility crisis, the fundamental problem is that the results of many peer-reviewed papers and reports have not been confirmed when the experiments have been repeated or the data reanalysed. Eminent medical scientist John Ioannidis belled the cat as early as 2005 in a much cited technical paper, Why Most Published Research Findings are False.

He concluded that "there is increasing concern that most current published research findings are false … For many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias." He further noted research findings were less likely to be true when "more teams are involved in a scientific field in chase of statistical significance".

There is a variety of reasons for the failure of studies to be replicated. At one end of the spectrum is fraud and misconduct, while at the other end is manipulation and cherry-picking of data. Researchers have strong incentives to establish significant results while discarding inconvenient data and failed hypotheses. Authors often deliberately make it difficult for other researchers to re-do experiments or check findings.

Additionally, many referees, who are the gatekeepers in the peer review process, do a lousy job by simply reading papers and approving them if they agree with their findings. Peer review generally doesn't involve re-running experiments, for instance.

One editor of an academic journal was so troubled by the issue of non-reproducibility that he decided to send out already published papers to new reviewers for their assessments. Apart from the fact a reasonable proportion of reviewers didn't even recognise that the papers had already been published, several of the papers were actually rejected by the new reviewers. So much for the infallibility of peer review.

A serious effort was made in 2015 to replicate the findings of 100 experiments reported in three major psychology journals. Ninety-seven per cent of the original studies had reported significant results but only 36 per cent of the replicated studies could confirm these effects. This is a damning outcome.

More recently, a research project tried to reproduce 21 social science experiments published between 2010 and 2015 in the prestigious journals *Science* and *Nature*. Thirteen replication studies were successful, while eight others could find no effects at all.

The editors of *Nature* recently conducted a survey of nearly 1600 researchers. It was noted that 70 per cent of researchers had failed to reproduce other scientists' experiments. Ninety per cent of respondents felt reproducibility in science was a significant or slight crisis. Only 3 per cent thought it wasn't a crisis at all.

Whether economics should be regarded as a science is debatable, but a recent edition of the prominent *Economic Journal* included a reassessment of the results from several of its published papers. The conclusion drawn was that most of the underlying analyses were statistically underpowered, meaning no reliance could be placed on the conclusions. For the other studies that had enough power, there was a distinct tendency for the size of effects to be overstated.

A replication audit of 67 economics papers published in 13 prestigious journals was conducted by the US Federal Reserve and the Department of Treasury. Less than half the studies could be replicated, even with the help of the authors. "We assert that economics research is usually not replicable," concluded the authors.

If all this sounds alarming to the layperson, it should. After all, the results of many of these peer-reviewed studies have had practical effects, warning people to alter their diets or lifestyles as well as influencing public policy initiatives.

At this stage, the disciplines most under a cloud are social psychology, neuroscience, chemistry, medicine (including cancer biology) and economics. No doubt the list will continue to grow as more replication studies are undertaken, although this is often difficult as such studies are generally not government funded.

One of the main elements of this crisis as identified by Ioannidis is the tendency of researchers to dredge data to get the most significant results. (Nobel prize-winning economist Ronald Coase famously quipped: "If you torture the data long enough, it will confess to anything.") To this we can add the downplaying of any deficiencies in the underlying data.

In this context, it is interesting to note the findings of John McLean, an Australian who has been awarded a PhD for his audit of the HadCRUT4 data set on global temperatures used by the IPCC. There are a large number of anomalies in the data set. For instance, two stations tracking temperatures recorded monthly average figures above 80C. Another two stations in the Caribbean recorded averages of 0C. A station in Romania recorded minus 45C and there is data sourced from ships that are located 80km inland.

More worrying is the use by the IPCC of a small number of global temperature recordings from the 1860s and 1870s — coverage was about one-eighth of the world at that time — as the measure of pre-industrial temperature levels. The accuracy of this assumption is highly questionable.

When the British Met Office was asked to respond to these criticisms, the answer was along the lines that there was an awareness of these weaknesses but they were few in number and the Met Office continuously was working to improve the data set, and this would be available to the IPCC when it next produced a report. On the face of it, this looks like a very unsatisfactory response, particularly given what we know about the crisis in science more generally.

What does the replication crisis mean for the credibility of science? Should we trust science to reliably inform public policy decision-making? Or should we conclude the scientific world is basically a club of self-serving, like-minded individuals who do not welcome dissenting views and are sloppy to boot? Should we just forget about scientific research and go with our instincts?

In my view, the preferred middle course is along the following lines. All research findings should be treated cautiously. Journals and research outlets should sign up to an open and transparent code of conduct, and published authors should be made to release all the details of underlying experiments, the data sets and computer codes. Studies that find no effects should be considered for publication.

Research funding bodies should allocate a portion of their funding to replication studies. An urgent priority in Australia is for the replication of several contentious studies about the Great Barrier Reef in which the overseas authors have never been prepared to hand over the data or the codes.

There is no doubt science has an important role to play in our society and economy. But as University of California computational biologist Michael Eisen warns us: “We need to get away from the notion, proven wrong on a daily basis, that peer review of any kind at any journal means that the work of science is correct.” The leadership of the IPCC should take note.