Document 1.

Bad science: When 'breakthrough research' turns out to be fraudulent

It is in the nature of scientists to argue over the evidence for or against any important breakthrough. Sometimes announcements made in good faith do not stand up to detailed scrutiny, namely the replication of the research by other experts.

On other occasions, scientists can be duped by the misconduct of their own colleagues prepared to cherry-pick favourable data to suit their conclusions, or, even worse, to fabricate data and commit outright scientific fraud – the most heinous crime in science.

One of the best examples of fraudulent research in recent years was the work on the cloning of human embryos by the South Korean researcher Hwang Woo-Suk of Seoul National University who announced in two scientific studies published in 2004 and 2005 that he had isolated human embryonic stem cells.

It turned out that he had faked many of the results and that he had engaged in dubious ethical practices in obtaining the human eggs needed for the research. He was eventually charged and found guilty of embezzlement and bioethical violations.

Another case involving stem cells occurred in the past year with Haruko Obokata, a young cell biologist at the Riken research institute in Japan. Dr Obokata claimed, with her Japanese and American colleagues, to have created stem cells by bathing ordinary blood or skin cells in a weak acid solution.

She called the technique "stimulus-triggered acquisition of pluripotency" (STAP) and it promised to revolutionise medicine as it offered the hope of creating therapeutic stem cells from a patient's own skin or blood with a simple, cheap technique that could be performed in any well-equipped lab.

Unfortunately, it was shown that the scientific research paper contained errors and other scientists were unable to replicate the findings, leading to a complete retraction of the research. Dr Obokata, however, continues to believe that the technique works and is still trying to replicate here own findings.

Replication is of course at the heart of science. When chemists Stanley Pons and Martin Fleischmann announced in 1989 that they had achieved nuclear fusion at room temperatures – so-called "cold fusion" – physicists everywhere wanted to reproduce the findings. Nuclear fusion, which powers the Sun, was only thought to occur at extremely high temperatures. If it could occur at room temperatures it would open the door to cheap, unlimited and clean energy.

It was too good to be true because it turned out not to be true. No-one has been able to demonstrate cold fusion as described by Pons and Fleischmann

Sometimes a scientific announcement is made that chimes with a bigger philosophical significance. In 1996, for instance, NASA announced that it had found evidence of fossilised mini-microbes in a piece of a meteorite from Mars, which fell to Earth 13,000 years ago and was discovered in Antarctica in 1984.

The clear implication was that life had existed on Mars and that we on Earth were "not alone" in the Universe. "If this discovery is confirmed, it will surely be one of the most stunning

insights into our universe that science has ever uncovered," said President Bill Clinton on the day of the announcement on 7 August 1996.
The trouble, once again, was that the discovery could not be confirmed by other researchers. It may have been a bad day for the idea of extra-terrestrial life, but it was a triumph for the scientific method.

The Independent - Sunday 14 September 2014

Document 2.

Scientific method: Statistical errors

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

For a brief moment in 2010, Matt Motyl was on the brink of scientific glory: he had discovered that extremists quite literally see the world in black and white.
The results were "plain as day", recalls Motyl, a psychology PhD student at the University of Virginia in Charlottesville. Data from a study of nearly 2,000 people seemed to show that political moderates saw shades of grey more accurately than did either left-wing or right-wing extremists. "The hypothesis was sexy," he says, "and the data provided clear support." The *P* value, a common index for the strength of evidence, was 0.01 — usually interpreted as 'very significant'. Publication in a high-impact journal seemed within Motyl's grasp.
But then reality intervened. Sensitive to controversies over reproducibility, Motyl and his adviser, Brian Nosek, decided to replicate the study. With extra data, the *P* value came out as 0.59 —not even close to the conventional level of significance, 0.05. The effect had disappeared, and with it, Motyl's dreams of youthful fame[1].
It turned out that the problem was not in the data or in Motyl's analyses. It lay in the surprisingly slippery nature of the *P* value, which is neither as reliable nor as objective as most scientists assume. "*P* values are not doing their job, because they can't," says Stephen Ziliak, an economist at Roosevelt University in Chicago, Illinois, and a frequent critic of the way statistics are used.
For many scientists, this is especially worrying in light of the reproducibility concerns. In 2005, epidemiologist John Ioannidis of Stanford University in California suggested that most published findings are false; since then, a string of high-profile replication problems has forced scientists to rethink how they evaluate results.
At the same time, statisticians are looking for better ways of thinking about data, to help scientists to avoid missing important information or acting on false alarms.

Regina Nuzzo - 12 February 2014 – Nature (International weekly journal of science)

Document 3.

Peer review: a flawed process at the heart of science and journals

People have a great many fantasies about peer review, and one of the most powerful is that it is a highly objective, reliable, and consistent process. I regularly received letters from authors who were upset that the *BMJ*[1] rejected their paper and then published what they thought to be a much inferior paper on the same subject. Always they saw something underhand. They found it hard to accept that peer review is a subjective and, therefore, inconsistent process. But it is probably unreasonable to expect it to be objective and consistent. If I ask people to rank painters like Titian, Tintoretto, Bellini, Carpaccio, and Veronese, I would never expect them to come up with the same order. A scientific study submitted to a medical journal may not be as complex a work as a Tintoretto altarpiece, but it is complex. Inevitably people will take different views on its strengths, weaknesses, and importance.

So, the evidence is that if reviewers are asked to give an opinion on whether or not a paper should be published, they agree only slightly more than they would be expected to agree by chance.

Sometimes the inconsistency can be laughable. Here is an example of two reviewers commenting on the same papers.

> *Reviewer A:* *`I found this paper an extremely muddled paper with a large number of deficits'*
>
> *Reviewer B:* *`It is written in a clear style and would be understood by any reader'.*

This—perhaps inevitable—inconsistency can make peer review something of a lottery. You submit a study to a journal. It enters a system that is effectively a black box, and then a more or less sensible answer comes out at the other end. The black box is like the roulette wheel, and the prizes and the losses can be big. For an academic, publication in a major journal like *Nature* or *Cell* is to win the jackpot.

The evidence on whether there is bias in peer review against certain sorts of authors is conflicting, but there is strong evidence of bias against women in the process of awarding grants. The most famous piece of evidence on bias against authors comes from a study by DP Peters and SJ Ceci. They took 12 studies that came from prestigious institutions that had already been published in psychology journals. They retyped the papers, made minor changes to the titles, abstracts, and introductions but changed the authors' names and institutions. They invented institutions with names like the Tri-Valley Center for Human Potential. The papers were then resubmitted to the journals that had first published them. In only three cases did the journals realize that they had already published the paper, and eight of the remaining nine were rejected—not because of lack of originality but because of poor quality. Peters and Ceci concluded that this was evidence of bias against authors from less prestigious institutions.

This is known as the Mathew effect: `To those who have, shall be given; to those who have not shall be taken away even the little that they have'. I remember feeling the effect strongly when as a young editor I had to consider a paper submitted to the *BMJ* by Karl Popper. I was unimpressed and thought we should reject the paper. But we could not. The power of the name was too strong. So we published, and time has shown we were right to do so. The paper argued that we should pay much more attention to error in medicine about 20 years before many papers appeared arguing the same.

The editorial peer review process has been strongly biased against `negative studies', i.e. studies that find an intervention does not work. It is also clear that authors often do not even

bother to write up such studies. This matters because it biases the information base of medicine. It is easy to see why journals would be biased against negative studies. Journalistic values come into play. Who wants to read that a new treatment does not work? That's boring.
We became very conscious of this bias at the *BMJ;* we always tried to concentrate not on the results of a study we were considering but on the question it was asking. If the question is important and the answer valid, then it must not matter whether the answer is positive or negative. I fear, however, that bias is not so easily abolished and persists.

Richard Smith - Journal of the Royal Society of Medicine – April 2006

[1.] The British Medical Journal

Document 4.

peer review
noun
: a process by which a scholarly work (such as a paper or a research proposal) is checked by a group of experts in the same field to make sure it meets the necessary standards before it is published or accepted

Merriam-Webster dictionary

Document 5.

Sidney Harris