Why so much medical research is rot

If you trawl around looking for statistical relationships the chances are you will find some. Yet many researchers still assume such findings to be meaningful, and many readers are still all too ready to take their word for it.

People born under the astrological sign of Leo are 15% more likely to be admitted to hospital with gastric bleeding than those born under the other 11 signs. Sagittarians are 38% more likely than others to land up there because of a broken arm.

Those are the conclusions that many medical researchers would be forced to make from a set of data presented to the American Association for the Advancement of Science by Peter Austin of the Institute for Clinical Evaluative Sciences in Toronto. At least, they would be forced to draw them if they applied the lax statistical methods of their own work to the records of hospital admissions in Ontario, Canada, used by Austin.

Austin, of course, does not draw those conclusions. His point was to shock medical researchers into using better statistics, because the ones they routinely employ today run the risk of identifying relationships when, in fact, there are none. He also wanted to explain why so many health claims that look important when they are first made are not substantiated in later studies.

The confusion arises because each result is tested separately to see how likely, in statistical terms, it was to have happened by chance. If that likelihood is below a certain threshold, typically 5%, then the convention is that an effect is 'real'. And that is fine if only one hypothesis is being tested. But if, say, 20 are being tested at the same time, then on average one of them will be accepted as provisionally true, even though it is not.

In his own study, Austin tested 24 hypotheses, two for each astrological sign. He was looking for instances in which a certain sign 'caused' an increased risk of a particular ailment. The hypotheses about Leos' intestines and Sagittarians' arms were less than 5% likely to have come about by chance, satisfying the usual standards of proof of a relationship. However, when he modified his statistical methods to take into account the fact that he was testing 24 hypotheses, not one, the boundary of significance dropped dramatically. At that point, none of the astrological associations remained.

Unfortunately, many researchers looking for risk factors for diseases are not aware that they need to modify their statistics when they test mul-

This article was first published by the *Economist Newspaper Limited*, London, on 22 February 2007, and is reprinted with permission

COURTESY HUBBLE SPACE TELESCOPE

tiple hypotheses. The consequence of that mistake, as John Ioannidis of the University of Ioannina School of Medicine, in Greece, explained to the meeting, is that a lot of observational health studies – those that go trawling through databases, rather than relying on controlled experiments – cannot be reproduced by other researchers.

Previous work by Ioannidis, on six highly cited observational studies, showed that conclusions from five of them were later refuted. In the new work he presented to the meeting, he looked systematically at the causes of bias in such research and confirmed that the results of observational studies are likely to be completely correct only 20% of the time. If such a study tests many hypotheses, the likelihood its conclusions are correct may drop as low as one in 1,000 – and studies that appear to find larger effects are likely, in fact, simply to have more bias.

So, the next time a newspaper headline declares that something is bad for you, read the small print. If the scientists used the wrong statistical method, you may do just as well believing your horoscope.

Sagittarians are 38% more likely than others to land up in hospital because of a broken arm