The Economist

The Economist explains
Explaining the world, daily
Sponsored by

The Economist explains

How to debunk a study

Aug 11th 2015, 23:55 by Y.Y.

FAME, glory, and a modest academic salary can all be yours if you write a major study. You might even change the world, as Michael Kremer and Edward Miguel found. They were the authors of an economics paper published in 2004 which showed that giving children deworming tablets increases their school attendance. The study sparked further research into cost-effective interventions in developing countries, all of which has led to millions of children being dewormed every year. Fame can also be won by those who debunk major studies. Last month a team at the London School of Hygiene and Tropical Medicine tried to replicate that study of Messrs Kremer and Miguel, and found only "some evidence, with high risk of bias" for the original conclusions. Next came the headlines, then the counter-fame, and finally the accusations, back and forth: of researchers capitalising on fame and headlines. Spectators called it the "worm wars". As the dust settles, it seems that the core message—it is good to deworm children—was not, after all, debunked. So how does one go about debunking a study properly?

It is crucial to understand the process of discovering an important result. Humans have a useful but unreliable tendency to find patterns amid meaningless noise. Scientists use statistical tests to sniff out sense from the data, but even their tests can sometimes turn up apparent relationships where there are none. As one guard against researchers making a mountain out of a molehill, each of the statistical relationships that scientists publish in their papers comes with a "p-value" attached. This is the probability of a false-positive finding—the likelihood, if the scientist were to

run their chosen statistical test on random data with no underlying pattern, that the test would turn up positive regardless. A lower p-value is better, as this makes it less likely the pattern came about for no reason. A simple way of debunking any paper is to recalculate the results of the original tests (a "strict replication"), hoping to spot an error in either the original calculation of a correlation or the p-value.

Debunkers themselves must beware that they are not committing the opposite sin, of making a molehill out of a mountain. The "power", or sensitivity, of a test is the likelihood that it sniffs out a positive relationship when one is actually there. Higher power is better. One method the medical researchers used in their replication was to cut a large two-year sample of test subjects into two one-year samples. This reduces the sample size, which gives the test less information to go on. This, in turn, reduces the power of the test, meaning that a greater number of positive relationships can sneak by undetected. The original study's authors say this is why the re-analysis didn't pick up on the relationship between deworming and school attendance.

So which kind of statistical test should be used for a social-scientific study? Amazingly, there is no one right answer, particularly when multiple disciplines are involved and the data are messy. The medical researchers defended their choice of method as reflecting the common practice in medicine, where tightly controlled and randomised trials are the norm. But methods appropriate to medicine may be too stringent in other contexts, and thus overly dismissive of positive results discovered by economists. Replicators should be forthright about the power of their tests, as much as correlation-hunters must disclose their p-values. The first problem, though, if you wish to debunk a study, is that data are not usually shared, and so replication cannot be done at all. In this instance, the original authors had taken the brave and unusual step of making their data widely available. Most studies never receive a fragment of the scrutiny that was placed on theirs. Chris Blattman, a professor at Columbia University, urges caution all round: "We should remember that most scientific studies don't stand up to scrutiny very well, and most are utterly wrong." The whole of the scientific world is ripe for debunking.