

OPINION

Power failure

BY VIRGINIA HUGHES

17 MAY 2013





Say you're poring over a brain imaging study that's making a hot new claim. If it has the power of an average neuroscience study, the odds that its findings are actually true are **50-50 or lower**, according to a report in the May *Nature Reviews Neuroscience*.

A study sample, by definition, is a slice of reality. With the proper statistical number crunching, researchers can quantify the likelihood that a pattern seen in a sample occurs in the real world.

Statistical power — the probability that a study will find an effect if one actually exists — depends on two factors: sample size and **effect size**.

In a brain imaging study that's looking for a telltale autism signature, for example, the sample size is the total number of participants, and the effect size is the difference in brain activity between the autism and control groups.

If a study has a small sample size, it needs to find a moderate-to-large effect in order to **surpass the power threshold** for statistical significance.

What surprised me, however, is that even when a small study finds a statistically significant effect, it's more likely to be false than a finding from a large study.

That's partly because random fluctuations between individual data points carry more weight in a small sample than they would in a large one. This works in both directions: Small studies sometimes underestimate the true size of the effect and sometimes overestimate it.

The trouble is, the overestimates are much more likely to be published in the scientific literature. What's more, replication studies that might catch the error rarely make it to top-tier journals.

The new report calculated the number of low-powered studies in neuroscience. Analyzing data from 48 meta-analyses published in 2011, **Kate Button** and her colleagues at the University of Bristol in the U.K. found that in most studies, the power ranges from just 8 percent to 31 percent. At this low power, a positive result is true only about half the time or less.

The researchers also tried to pinpoint the types of studies most vulnerable to small sample size. It turns out that studies measuring brain volume have a statistical power of about eight percent. Research on animals, too, is often severely underpowered for the effects researchers claim to find.

None of the studies in this analysis were focused on autism, Button says. "But I would expect the problems associated with low power would apply to the autism literature — they seem to be fairly ubiquitous across science disciplines."

I see two lessons here for scientists. One is that you should always try to increase sample sizes, perhaps by working collaboratively with other teams, and by thoroughly describing your methods so others can easily reproduce them.

The second lesson pertains not just to scientists, but to students, journalists and the public at large: Be skeptical of everything you read, even — and maybe especially — if the source is a prestigious scientific journal.