

Scholars take aim at false positives in research

 news.uchicago.edu/story/scholars-take-aim-false-positives-research

Prof. John List is co-author of a new paper recommending changes to the p-value standard.

By [Thomas Gaulkin](#)

Sep 1, 2017

UChicago professor argues for lowering key statistical benchmark

A single change to a century-old statistical standard would dramatically improve the quality of research in many scientific fields, shrinking the number of so-called false positives, [according to a commentary](#) published Sept. 1 in *Nature Human Behaviour*.

The argument, co-authored by University of Chicago economist John List, represents the consensus of 72 scholars from institutions throughout the world and disciplines ranging from neurobiology to philosophy. Their recommendations could have a major effect on the publication of academic work and on public policy.

“We advertise interventions as working because statistically we think they’re working. But they’re actually not working. This is becoming a crisis in the sciences,” said List, the Kenneth C. Griffin Distinguished Service Professor in Economics.

List and his co-authors suggest that scientists need to reset a statistical benchmark known as the p-value because the standards of evidence for claiming new discoveries in many fields are simply too low. The approach is damaging to the credibility of scientific claims, they said.

A p-value standard was adopted beginning in the 1920s, when British statistician Ronald Fisher proposed a value below 0.05 as a threshold to determine the validity of research findings. If the p-value falls below that threshold—meaning the probability that a study’s conclusions are due to random chance is below 5 percent—then the research is generally considered to be statistically significant.

But the p-value threshold has become a target of criticism in response to a perceived replication crisis in scientific communities. Science journals frequently use statistical significance—and p-values—as a test for selecting which papers to publish. List said the current p-value threshold of 0.05 is allowing many studies to be published and influence economic and political decisions even though the results may not be reproducible by other researchers.

“If Ronald Fisher would have known that close to a 100 years later we would be using the 0.05 standard religiously to make ‘informed’ policy decisions, I don’t think he would have advanced

it," List said.

More reproducible studies

To be sure that an initial discovery will work when put into practice, results should be replicable. Previous studies have shown that only 24 percent of psychology studies with a p-value of 0.05 could be confirmed by further experiments, suggesting that three out of four studies presented false positive results. Similarly, only 44 percent of economics papers with the same p-value were reproducible.

The authors calculated that lowering the p-value threshold to 0.005 would roughly double rates of replication in psychology and economics, and other fields would see similar outcomes. "Changing the p-value threshold is simple, aligns with the training undertaken by many researchers and might quickly achieve broad acceptance," the authors said.

List agrees. "You want to set up a world where you have more people trying to replicate, and you want society to reward those people," he said. "And you also want more results that go into policy to be true results, to be replicable. Under the 0.005 more of them would be."

To further encourage publication and replication of studies, the authors of the paper propose that new findings that currently would be called "significant" but don't meet the revised 0.005 p-value should be called "suggestive" instead.

List and his co-authors are careful to point out that a change to the p-value is not the only step to improve scientific research. "We have diverse views about how best to improve reproducibility, and many of us believe that other ways of summarizing the data...are preferable to p-values," they said.