Over the last several years, psychological scientists have become especially concerned about the reproducibility of studies in the field. Do peer-reviewed publications hold up under scientific scrutiny? Or are some papers that get published just lucky flukes?

Until recently, researchers have relied only on intuition to estimate reproducibility. A new report published in *Science*, however, attempts to provide the first empirical estimate of the reproducibility of psychological science. According to this report, less than half of the psychology studies from a sample of 100 replicated.

The report, coordinated by APS Fellow Brian Nosek (University of Virginia) and the Center for Open Science in Charlottesville, VA, involved recruiting over 270 researchers who attempted to reproduce 100 findings published in psychology journals in 2008.

Just because a study was not replicated does not mean it was wrong, however. Replication failures can sometimes occur when the replication misses detecting a real effect or when the methodology of the replication differs in important ways from the methodology of the original study.

Among the 100 studies selected for the replication project were 40 published in APS’s flagship journal, *Psychological Science*. Replication teams worked with the authors of the original studies when possible and posted their data and analyses online for public evaluation. The set of replications took over 3 years to complete.

The replication teams’ findings were striking: Overall, 97% of the original studies reported statistically significant \( p \) values below .05, but only 36% of the replication studies found statistically significant results (\( p < .05 \)). Moreover, whereas the effect sizes in the original studies were moderate, on average —
Pearson $r = .40$ — in the replications, the sizes of the effects were $r = .20$ — half as large as the originals.

Nosek and colleagues also assessed differences within subfields of psychology. Cognitive psychology studies were twice as likely to replicate as were social psychology studies, but both subfields showed equivalent decreases in effect sizes in the replication attempts. The researchers also searched for factors associated with whether a replication attempt succeeded or failed. Success was related to the original strength of evidence, but not to factors such as the experience or expertise of the replication team.

According to Nosek, many studies fail to reproduce because scientists are rewarded for getting research published, and some findings are simply more likely to be accepted for publication.

“I am more likely to get published for a positive result than a negative, with a novel result than a registered replication, and with a very clean story, as opposed to one with lots of loose ends,” he stated at a recent presentation at the National Science Foundation. “Because we’re incentivized to make it a novel, positive, clean story, then, there’s lots of reasons for me and for my individual success to find ways to make it as beautiful as possible, even if that makes it look a lot different from what the actual evidence is.”

The project findings probably mean that psychological science needs to devote more attention to improving reproducibility, Nosek emphasized in a teleconference announcing the results of the report.

“But, I don’t see this story as pessimistic,” he added. “The project is a demonstration of science demonstrating one of its central qualities — self-correction.”

Indeed, APS has been leading the way in encouraging self-correction in psychological science, APS Executive Director Alan Kraut commented in the same teleconference.

“We have changed how articles are published in *Psychological Science*, changes that encourage greater transparency and stronger statistical analyses and that provide special recognition for pre-registering hypotheses and for sharing materials and data,” he said. “APS also is pushing at the leading edge on issues of replicability.”

The badge program recognizing open science practices, *Registered Replication Reports*, and the *Transparency and Openness Guidelines*, of which APS is a signatory, are three examples of APS’s efforts in this arena. These programs are likely to lead to an improvement in the reproducibility of psychological science, said Interim Editor of *Psychological Science* D. Stephen Lindsay in a statement.

“It is exciting to anticipate a future replication of this extraordinary project in, say, 8 years, testing the replicability of articles published in *Psychological Science* in 2016; if we do our jobs correctly then the replication rate will be dramatically higher,” he said.

“Replication is a fundamental part of science — it is science at its best,” Kraut echoed.