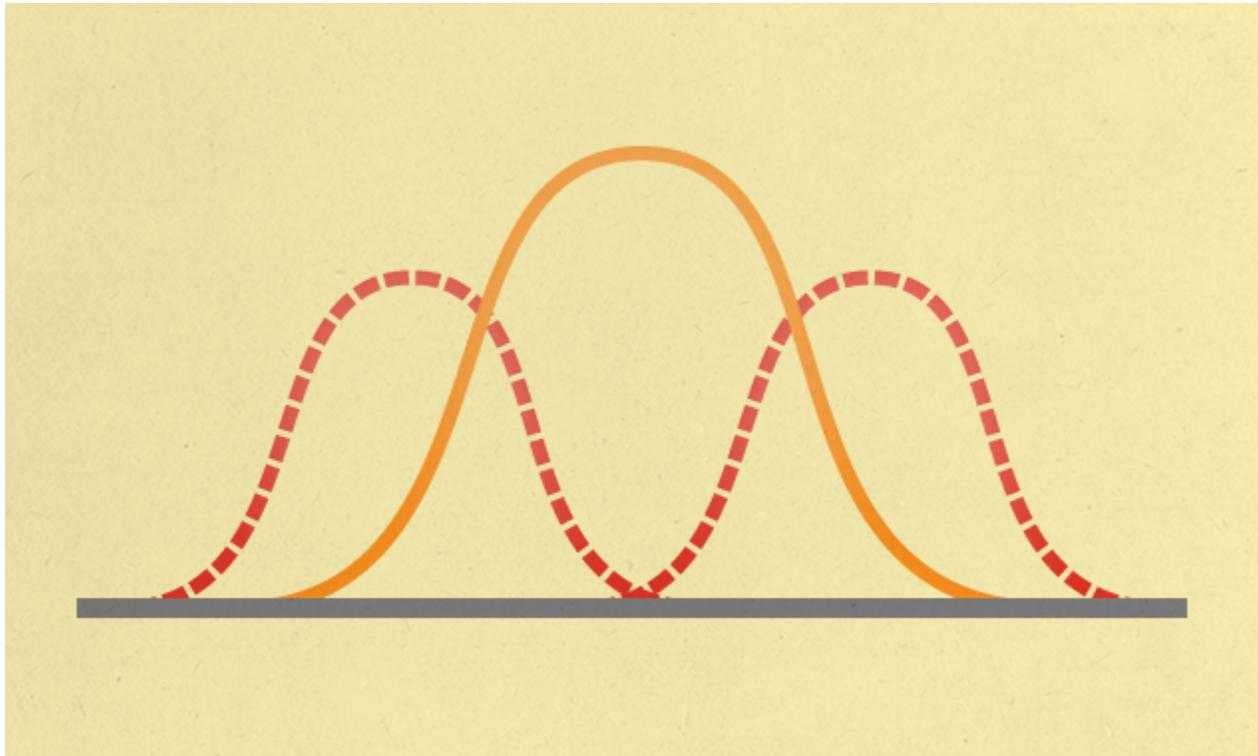


## 'One-size-fits-all' threshold for $P$ values under fire

Scientists hit back at a proposal to make it tougher to call findings statistically significant.

Dalmeet Singh Chawla

19 September 2017



Researchers are at odds over when to dub a discovery 'significant'. In July, 72 researchers took aim at the  $P$  value, calling for a lower threshold for the popular but much-maligned statistic. In a response published on 18 September<sup>1</sup>, a group of 88 researchers have responded, saying that a better solution would be to make academics justify their use of specific  $P$  values, rather than adopt another arbitrary threshold.

$P$  values have been used as measures of significance for decades, but academics have become increasingly aware of their shortcomings and the potential for abuse. In 2015, one psychology journal banned  $P$  values entirely.

The statistic is used to test a ‘null hypothesis’, a default state positing that there is no relationship between the phenomena being measured. The smaller the  $P$  value, the less likely it is that the results are due to chance — presuming that the null hypothesis is true. Results have typically been deemed ‘statistically significant’ — and the null hypothesis dismissed — when  $P$  values are below 0.05.

In a July preprint, since published in *Nature Human Behaviour*<sup>2</sup>, researchers, including leaders in the push for greater reproducibility, said that this threshold should be reduced to 0.005 to keep false positives from creeping into social sciences and biomedical literature.

But “setting this one threshold for all sciences is too extreme,” says Daniel Lakens, an experimental psychologist at Eindhoven

University of Technology in the Netherlands and lead author of the new commentary, which was posted to the PsyArXiv preprint server. “The moment you ask people to justify what they are doing, science will improve,” he adds.

#### Related stories

- Big names in statistics want to shake up much-maligned P value
- Statisticians issue warning over misuse of P values
- Reproducibility: A tragedy of errors

**More related stories**

#### Related stories

- Big names in statistics want to shake up much-maligned P value
- Statisticians issue warning over misuse of P values
- Reproducibility: A tragedy of errors

**More related stories**

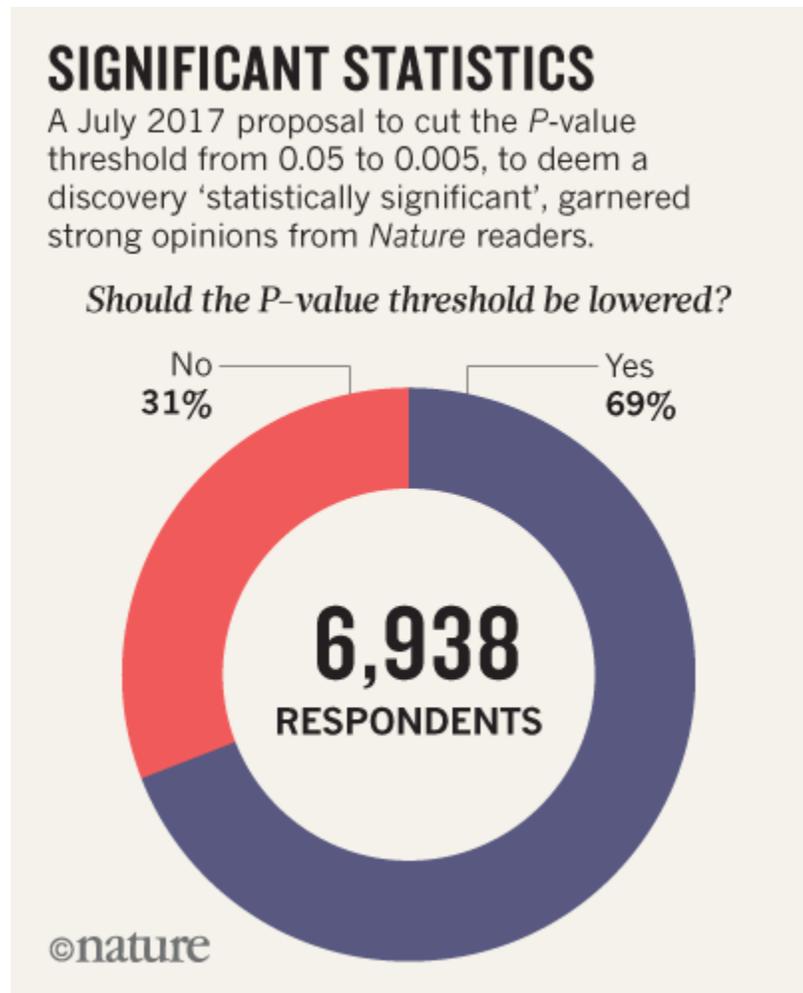
## Unintended consequences

Some researchers worry that lowering  $P$  value cut-offs may exacerbate the 'file-drawer problem', when studies containing negative results are left unpublished. A more stringent  $P$  value threshold could also lead to more false negatives — claiming that an effect doesn't exist when in fact it does. "Before you implement any policy, you want to be more certain that there are no unintended negative consequences," says Lakens.

Instead, Lakens and colleagues say, researchers should select and justify  $P$  value thresholds for their experiments, before collecting any data. These levels would be based on factors such as the potential impact of a discovery, or how surprising it would be. Such thresholds could then be evaluated via their registered reports, a type of scientific article in which methods and proposed analyses are peer-reviewed before any experiments are conducted.

"I don't think researchers will ever have an incentive to say they need to use a more stringent threshold of evidence," counters Valen Johnson, a statistician at Texas A&M University in College Station who is a co-author of the July manuscript. And many scientists are likely to go easy on their own work, says another co-author, Daniel Benjamin, a behavioural economist at the University of Southern California, Los Angeles.

But Lakens thinks that any attempts to manipulate  $P$  values will be obvious from the justifications that researchers pick. "At least everyone agrees that it's good to change the mindless use of 0.05," he says.



Setting specific thresholds for standards of evidence is “bad for science”, says Ronald Wasserstein, executive director of the American Statistical Association, which last year took the unusual step of releasing explicit recommendations on the use of *P* values for the first time in its 177-year history. Next month, the society will hold a symposium on statistical inference, which follows on from its recommendations.

Wasserstein says he hasn't yet taken a position on the current debate over *P* value thresholds, but adds that “we shouldn't be surprised that there isn't a single magic number”.

*Nature* doi:10.1038/nature.2017.22625

## References

---

1. Lakens, D. *et al.* <https://psyarxiv.com/9s3y6> (2017).
2. Benjamin, D. J. *et al.* *Nature Hum. Behav.* <http://dx.doi.org/10.1038/s41562-017-0189-z> (2017).