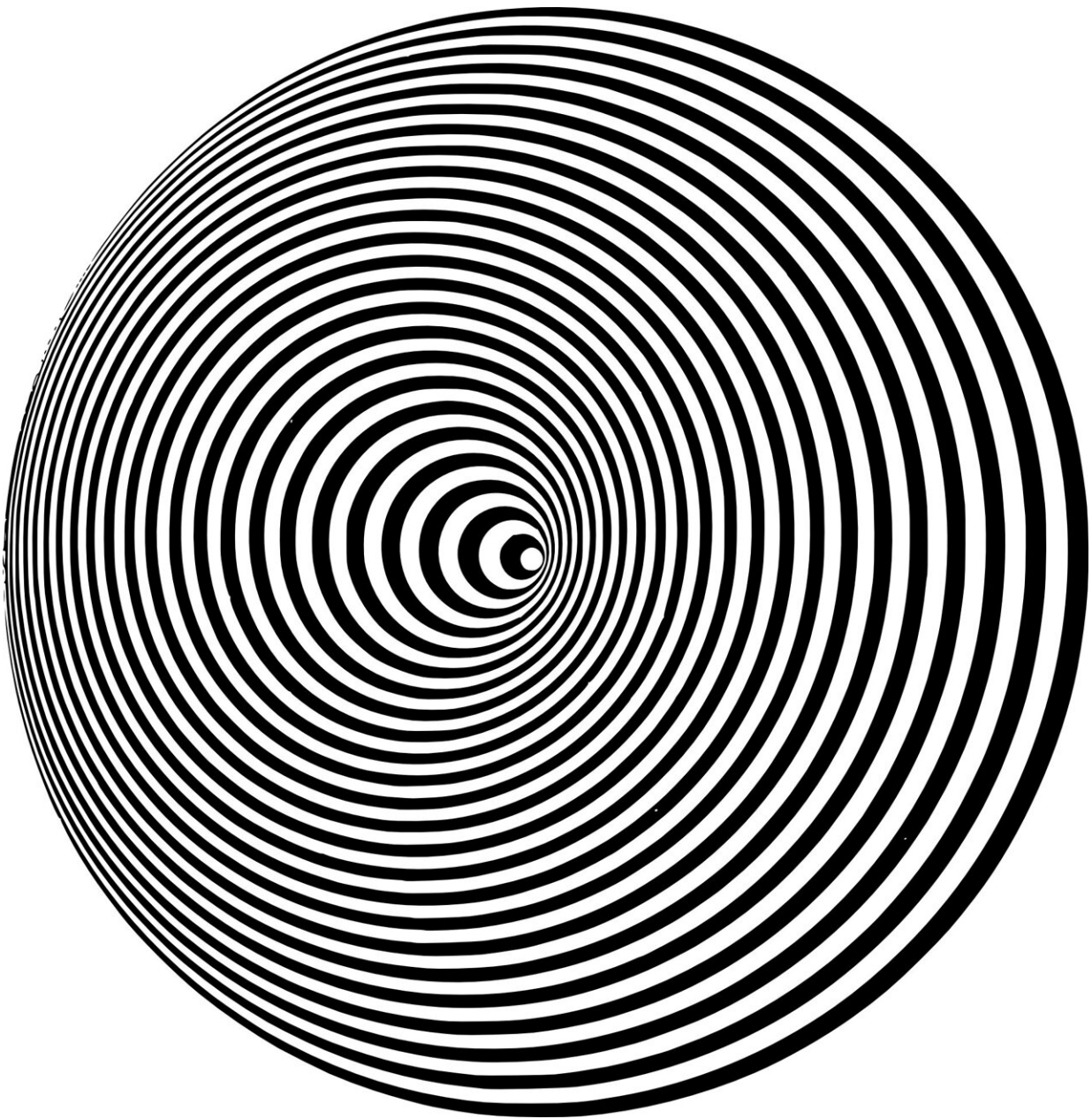


# Time to abandon null hypothesis significance testing? Moving beyond the default approach

December 13 2023

---



Credit: Pixabay/CC0 Public Domain

Researchers from Northwestern University, University of Pennsylvania, and University of Colorado published a [new \*Journal of Marketing\* study](#) that proposes abandoning null hypothesis significance testing (NHST) as the default approach to statistical analysis and reporting.

The study is titled "'Statistical Significance' and Statistical Reporting: Moving Beyond Binary" and is authored by Blakeley B. McShane, Eric T. Bradlow, John G. Lynch, Jr., and Robert J. Meyer.

Null hypothesis significance testing (NHST) is the default approach to [statistical analysis](#) and reporting in marketing and, more broadly, in the biomedical and social sciences. As practiced, NHST involves

1. assuming that the intervention under investigation has no effect along with other assumptions,
2. computing a statistical measure known as a *P*-value based on these assumptions, and
3. comparing the computed *P*-value to the arbitrary threshold value of 0.05.

If the *P*-value is less than 0.05, the effect is declared "statistically significant," the assumption of no effect is rejected, and it is concluded that the intervention has an effect in the [real world](#). If the *P*-value is above 0.05, the effect is declared "statistically nonsignificant," the assumption of no effect is not rejected, and it is concluded that the intervention has no effect in the real world.

## Criticisms of NHST

Despite its default role, NHST has long been criticized by both statisticians and applied researchers, including those within marketing. The most prominent criticisms relate to the dichotomization of results into "statistically significant" and "statistically nonsignificant."

For example, authors, editors, and reviewers use "statistical (non)significance" as a filter to select which results to publish. Meyer says that "this creates a distorted literature because the effects of published interventions are biased upward in magnitude. It also encourages harmful research practices that yield results that attain so-called statistical significance."

Lynch adds that "NHST has no basis because no intervention has precisely zero effect in the real world and small  $P$ -values and '[statistical significance](#)' are guaranteed with sufficient sample sizes. Put differently, there is no need to reject a hypothesis of zero effect when it is already known to be false."

Perhaps the most widespread abuse of statistics is to ascertain where some statistical measure such as a  $P$ -value stands relative to 0.05 and take it as a basis to declare "statistical (non)significance" and to make general and certain conclusions from a single study.

"Single studies are never definitive and thus can never demonstrate an effect or no effect. The aim of studies should be to report results in an unfiltered manner so that they can later be used to make more general conclusions based on cumulative evidence from multiple studies. NHST leads researchers to wrongly make general and certain conclusions and to wrongly filter results," says Bradlow.

" $P$ -values naturally vary a great deal from study to study," explains McShane. As an example, a "statistically significant" original study with an observed  $P$ -value of  $p = 0.005$  (far below the 0.05 threshold) and a

"statistically nonsignificant" replication study with an observed  $P$ -value of  $p = 0.194$  (far above the 0.05 threshold) are highly compatible with one another in the sense that the observed  $P$ -value, assuming no difference between them, is  $p = 0.289$ .

He adds that "however when viewed through the lens of 'statistical (non)significance,' these two studies appear categorically different and are thus in contradiction because they are categorized differently."

## Recommended changes to statistical analysis

The authors propose a major transition in statistical analysis and reporting. Specifically, they propose abandoning NHST—and the  $P$ -value thresholds intrinsic to it—as the default approach to statistical analysis and reporting. Their recommendations are as follows:

- "Statistical (non)significance" should never be used as a basis to make general and certain conclusions.
- "Statistical (non)significance" should also never be used as a filter to select which results to publish.
- Instead, all studies should be published in some form or another.
- Reporting should focus on quantifying study results via point and interval estimates. All of the values inside conventional interval estimates are at least reasonably compatible with the data given all of the assumptions used to compute them; therefore, it makes no sense to single out a specific value, such as the null value.
- General conclusions should be made based on the cumulative evidence from multiple studies.
- Studies need to treat  $P$ -values continuously and as just one factor among many—including prior evidence, the plausibility of mechanism, study design, [data quality](#), and others that vary by research domain—that require joint consideration and holistic integration.

- Researchers must also respect the fact that such conclusions are necessarily tentative and subject to revision as new studies are conducted.

Decisions are seldom necessary in scientific reporting and are best left to end-users such as managers and clinicians when necessary.

In such cases, they should be made using a decision analysis that integrates the costs, benefits, and probabilities of all possible consequences via a loss function (which typically varies dramatically across stakeholders)—not via arbitrary thresholds applied to statistical summaries such as *P*-values ("statistical (non)significance") which, outside of certain specialized applications such as industrial quality control, are insufficient for this purpose.

**More information:** Blakeley B. McShane et al, EXPRESS: "Statistical Significance" and Statistical Reporting: Moving Beyond Binary, *Journal of Marketing* (2023). [DOI: 10.1177/00222429231216910](https://doi.org/10.1177/00222429231216910)

Provided by American Marketing Association

Citation: Time to abandon null hypothesis significance testing? Moving beyond the default approach (2023, December 13) retrieved 28 April 2024 from <https://phys.org/news/2023-12-abandon-null-hypothesis-significance-default.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.