

# How Marginal Are 'Marginally Significant' $p$ -Values?

February 27, 2019

As the research community debates whether the  $p$ -value should be swept into the statistical dustbin, the question remains: How are authors actually presenting  $p$ -values? Are authors reporting only the values that make the .05 cutoff or are they reporting every  $p$ -value, significant or not? And for the values that reside above .05, how often do authors succumb to the temptation of the “marginally significant”?

In a 2016 study in *Psychological Science*, Pritschet and colleagues found cause for concern, showing an increase in the number of articles containing marginally significant results reported over time. But Tilburg University researchers **Anton Olsson-Collentine**, **Marcel A. L. M. van Assen**, and **Chris H. J. Hartgerink** found a different trend when they accounted for base rates. These findings appear in *Psychological Science*.

Olsson-Collentine and colleagues argue that since authors now report more  $p$ -values per article than they used to, more articles will also contain  $p$ -values between .05 and .10. Consequently, even if the proportion of  $p$ -values reported as “marginally significant” stays the same over time, one would expect more articles to contain “marginally significant” results. In other words, observing that more articles contain “marginally significant” results doesn’t necessarily mean that the tendency to report any given  $p$ -value as “marginally significant” is actually increasing.

The researchers used regular expressions to search for and automatically extract  $p$ -values from articles published in 70 American Psychological Association journals between 1985 and 2016.

Searching for any mention of “margin\*” and “approach\*” in the 200 characters preceding and succeeding any  $p$ -value result, the researchers obtained a final sample of 42,504  $p$ -values between .05 and .10.

In line with results reported in Pritschet et al. in 2016, the results showed an increase in articles reporting “marginally significant” results in the two journals, *Journal of Personality* and *Social Psychology and Developmental Psychology*.

But closer inspection of the data revealed a more complex story. In *Developmental Psychology*, the percentage of  $p$ -values between .05 and .10 that were described as “marginally significant” actually decreased over time, but this was masked by an increase in the overall number of reported  $p$ -values that fell between .05 and .10.

This finding “demonstrates the importance of distinguishing results at the level of the articles from those at the level of  $p$ -values,” Olsson-Collentine and colleagues write.

Overall, the researchers found results described as “marginally significant” to be quite common, characterizing about 40% of all the  $p$ -values in the sample that fell between .05 and .10. Across nine

psychology disciplines represented in the journals, they found the practice to be most common in journals focused on organizational psychology (45% of  $p$ -values between .05 and .10) and least common in those focused on clinical psychology (30% of  $p$ -values between .05 and .10).

Of note, the results showed that the percentage of  $p$ -values reported as “marginally significant” decreased over time across all journals, and also within most of the disciplines. In no discipline was there evidence of an increasing percentage of “marginally significant” results, although the trend was largely stable over time for several disciplines.

Olsson-Collentine, van Assen, and Hartgerink suggest several possible explanations for decreasing usage of “marginally significant” to describe individual  $p$ -values, including increasing statistical awareness on the part of researchers and increasingly stringent editorial criteria.

“Such a high prevalence is a call for disciplines and journal editors to examine where they stand on the reporting of  $p$ -values as marginally significant,” Olsson-Collentine says. “We recommend not interpreting  $p$ -values between .05 and .1 as marginally significant due to their low evidential value, and note that doing so might be an indication of post-hoc flexibility in decision rules.”

## References

Olsson-Collentine, A., van Assen, M. A. L. M., & Hartgerink, C. H. J. (2019). The prevalence of marginally significant results in psychology over time. *Psychological Science*.

[doi.org/10.1177/0956797619830326](https://doi.org/10.1177/0956797619830326)

Pritschet, L., Powell, D., Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science*, 27(7),

1036–1042. [doi.org/10.1177/0956797616645672](https://doi.org/10.1177/0956797616645672)