

Rise in Reporting p -Values as “Marginally Significant”

May 20, 2016

A researcher collects data, runs a statistical test, and finds that the p value is approximately .07. What happens next? According to [a study](#) conducted by Laura Pritschet (University of Illinois at Urbana-Champaign), Derek Powell (University of California, Los Angeles), and Zachary Horne (also at the University of Illinois), that researcher may be likely to report that result as “marginally significant” — not quite significant, but getting there. While it may be common, Pritschet and colleagues argue that this practice is “rooted in serious statistical misconceptions” and is likely to lead to false-positive errors (and sometimes false negatives, too). To make matters worse, evidence suggests that this practice is on the rise.

Pritschet, Powell, and Horne note that the practice of reporting marginally-significant results is problematic for two main reasons. First, the field of psychological science has no agreed-upon standards for how and when results should be reported as marginally significant. The second edition of the American Psychological Association style manual, published in 1974, advised, “Do not infer trends from data that fail by a small margin to reach the usual levels of significance.” This language was soon cut, however, and information about marginal significance has been absent from the manual for over 30 years.

Potentially more troubling is that reporting of marginally-significant results mixes two types of scientific reasoning: Neyman-Pearson decision theory, which relies on hard cutoffs, and Fisher’s hypothesis testing approach, in which a p value can be considered a measure of evidence.

“The concept of marginal significance is dubious under either framework,” warn the authors.

To investigate reporting of marginal results, Pritschet and colleagues examined the papers published each decade between 1970 and 2010 in three top cognitive, developmental, and social psychology journals published every decade — 1,535 in all. The authors examined each article for the reporting of a marginally-significant result, by searching for the terms “margin” and “approach,” and examined the first instance of a marginal report when it appeared.

Pritschet and colleagues found that researchers called many different p values marginal, ranging from $p = .05$ to $p = .18$. Over 90% of p values called marginal were between .05 and .10, however.

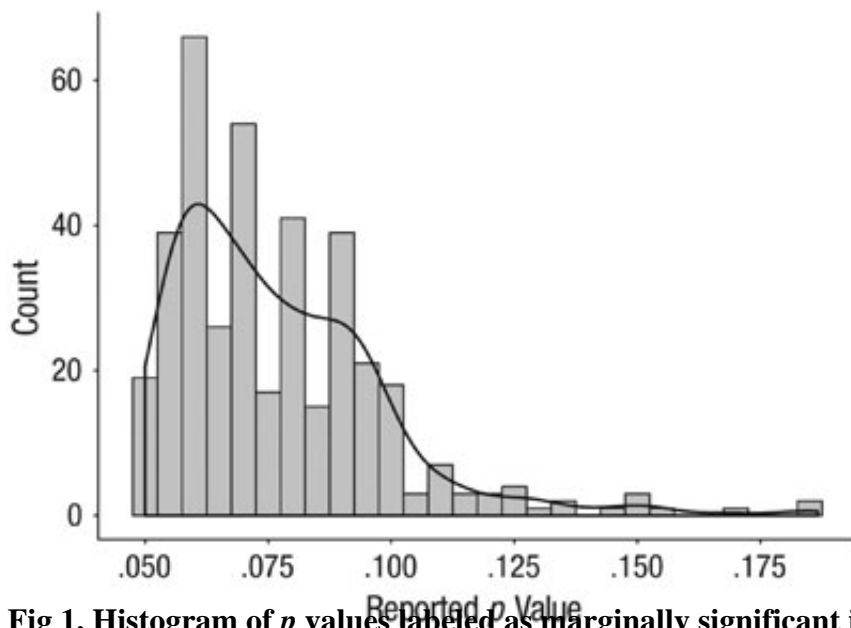


Fig 1. Histogram of p values labeled as marginally significant in the articles analyzed

Next, the researchers examined the proportion of articles that contained reporting of marginal reports, and how those proportions have changed over time. All journals showed an increase in reporting of marginal results: In 1970, 18% of articles examined described a p value as marginally significant, but in 2000, over half of all articles did so. The researchers noticed, too, that the social psychology journal was most likely to contain reporting of marginally-significant results.

In additional analyses, the researchers found that papers that reported results of more experiments were more likely to include reporting of a marginally-significant result. Even still, and controlling for this effect, the articles in the social psychology journal were more likely to report results of this type.

“We are reluctant to draw firm conclusions on the basis of these results alone,” the authors cautioned.

Pritschet and colleagues admit that the wider interpretation of these results is up for debate.

“Is the increased acceptance of marginally significant effects representative of a graded, Fisherian interpretation of p values, according to which hard cutoffs are thought to be arbitrary? Or might it suggest the emergence of a more questionable state of affairs for psychological methodology?” they wrote.

Either way, the authors advise a careful look at the reporting of marginal results within psychological science.

Reference

Pritschet, L., Powell, D., Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science*.
doi: <http://pss.sagepub.com/content/early/2016/05/14/0956797616645672.abstract>