

Published in *Psychological Reports*, 80 (1997), 337-338.

Publication Bias Against Null Results

Raymond Hubbard

College of Business and Public Administration, Duke University

J. Scott Armstrong

The Wharton School, University of Pennsylvania

Summary

Studies suggest a bias against the publication of null ($p > .05$) results. Instead of significance, we advocate reporting effect sizes and confidence intervals, and using replication studies. If statistical tests are used, power tests should accompany them.

A number of studies have shown that peer review is biased against the publication of null ($p > .05$) or so-called “negative,” results. Sterling, Rosenbaum, and Weinkam’s (1995) recent article cited ten studies on this. They also conducted a further analysis showing that 95.6% of psychology papers using tests of significance rejected the null hypothesis, while for medical journals it was 85.4%. We briefly mention the results of four additional empirical studies on the incidence of publication of articles rejecting the null hypothesis.

Using Sterling’s classification procedures, Wilson, Smoke, and Martin’s (1973) content analysis of the *American Journal of Sociology*, *American Sociological Review*, and *Social Forces* for the period July 1969 to June 1970 indicated that 61 of 76 (80.3%) articles using tests of significance rejected the null hypothesis. Greenwald’s (1975) data came from a single annual (1972) issue of the *Journal of Personality and Social Psychology*. He found that 175 of 199 (87.9%) articles managed to reject the null hypothesis. Lindsay (1994), also employing Sterling’s definitions, surveyed all empirical budgeting and control articles published in *Accounting*, *Organizations and Society*, *The Accounting Review*, and *Journal of Accounting Research* during the 1970-1987 period; 84.2% of the 38 studies rejected the null hypothesis. Again, using the methodology proposed by Sterling, et al., Hubbard and Armstrong (1992) analyzed 32 randomly selected issues of each of the *Journal of Marketing*, *Journal of Marketing Research*, and *Journal of Consumer Research* for the period 1974 through 1989. Of the 692 papers using tests of significance, 638 (92.2%) rejected H_0 .

Statistical significance testing is the entrenched method for evaluating the merits of hypotheses in psychology and elsewhere despite increasing concerns about its usefulness (Bakan, 1966; Cohen, 1994). Indeed, the American Psychological Association’s Board of Scientific Affairs is considering the possibility of phasing out such testing in textbooks and journal articles (Shea, 1996).

There are several alternatives to significance testing. One approach is to measure effect sizes and associated confidence intervals. Another is to conduct replications and extensions, as these provide more appropriate ways to assess the confidence one might have in research findings (Hubbard & Armstrong, 1994; Hubbard & Vetter, 1996)

Those who persist in using significance tests should report the level of significance and perform a statistical power analysis. (The power of a test is the probability of rejecting a false null hypothesis.) This applies especially to null results. Statistically non-significant findings accompanied by high power can be contributions to knowledge. They can provide evidence of a trivial effect size in the population, reduce sampling bias in meta-analyses, and prevent researchers from reinvestigating blind alleys.

Using procedures outlined by Cohen (1988), Hubbard and Armstrong (1992) examined the statistical power of the null outcomes found in their study. The average power levels were high; the probabilities of detecting small, medium, and large effect sizes in the population were .35, .89, and .99, respectively. Other things being equal, null results with adequate power are likely to be worthy of publication.

References

- Bakan, D. (1966), "The test of significance in psychological research," *Psychological Bulletin*, 66, 1-29.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1994), "The earth is round ($p < .05$)," *American Psychologist*, 49, 997-1003.
- Greenwald, A. G. (1975), "Consequences of prejudice against the null hypothesis," *Psychological Bulletin*, 82, 1-20.
- Hubbard, R., & Armstrong, J. S. (1992), "Are null results becoming an endangered species in marketing?" *Marketing Letters*, 3, 127-136.
- Hubbard, R., & Armstrong, J. S. (1994), "Replications and extensions in marketing: Rarely published but quite contrary," *International Journal of Research in Marketing*, 11, 233-248.
- Hubbard, R., & Vetter, D. E. (1996), "An empirical comparison of published replication research in accounting, economics, finance, management, and marketing," *Journal of Business Research*, 35, 153-164.
- Lindsay, R. M. (1994), "Publication system biases associated with the statistical testing paradigm," *Contemporary Accounting Research*, 11, 33-57.
- Shea, C. (1996), "Psychologists debate accuracy of 'significance test.'" *The Chronicle of Higher Education*, 49 (August 16), A12 & A17.
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995), "Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa," *The American Statistician*, 49, 108-112.
- Wilson, F. D., Smoke, G. L., & Martin, J. D. (1973), "The replication problem in sociology: A report and a suggestion," *Sociological Inquiry*, 43, 141-149.