Communication Research Reports
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rcrr20

A Defense of Publishing Nonsignificant (ns) Results

Timothy R. Levine [a]
[a] School of Media and Communication , Korea University , Seoul , Republic of Korea
Published online: 12 Jul 2013.

Communication Research Reports
Vol. 30, No. 3, July–September 2013, pp. 270–274

Routledge
Taylor & Francis Group

SPOTLIGHT ON METHOD/ANALYSIS

A Defense of Publishing Nonsignificant (*ns*) Results

Timothy R. Levine

An argument is advanced for the propositions that a lack of statistically significant findings does not automatically justify rejecting a scientific article for publication or mean that the findings are necessarily uninformative. Systematically declining publication for nonsignificant results leads to negative consequences that include distorting scientific literatures, making hypotheses and theories less falsifiable, depriving meta-analyses of an accurate sample of prior findings, and encouraging questionable research practices. Providing effect sizes and confidence intervals can make findings more informative, regardless of whether the finding is $p < .05$.

Keywords: Falisibility; Meta-Analysis; Nonsignificant Results

> What a shame that this fine writing is wasted on nonsignificant [*ns*] finding after nonsignificant finding. I have occasionally used the phrase "null finding" to describe nonsignificance, but never before have I had to resort to the term "null manuscript." In my opinion, there is no way this manuscript should be published in CRR [*Communication Research Reports*]. I will spare the authors the "why-we-can't-publish-null-findings" lecture because they certainly know it full well. This was an admirable attempt to salvage the unsalvageable but, in my opinion, it didn't make it. (2012: anonymous reviewer)

Authors Levine, Shulman, Carpenter, DeAdrea, and Blair (2013) replied:

> We do not agree at all that $p > .05$ or that small effects are (a) in any way shameful, (b) uninformative, (c) or unworthy of publication. It is unfortunate that the

Timothy R. Levine (PhD, Michigan State University, 1992) is a professor in the School of Media and Communication at Korea University, Seoul, Republic of Korea. *Correspondence*: Timothy R. Levine, School of Media and Communication, 607 Media Hall, Korea University, Seoul, Republic of Korea; E-mail: levinet111@gmail.com

ISSN 0882-4096 (print)/ISSN 1746-4099 (online)
DOI: 10.1080/08824096.2013.806261

> reviewer did not include the argument, because we are curious what that might be. There are many good reasons why $p > .05$ is never a good reason to deny publication. We have never seen a valid argument otherwise. So, we are curious about what we were spared.

More than 40 years before this exchange, Lykken (1968) wrote the following:

> Statistical significance is perhaps the least important attribute of a good experiment; it is never a sufficient condition for claiming that a theory has been usefully corroborated, that a meaningful empirical fact has been established, or that an experimental report ought to be published. (p. 158)

Lykken's words were wise, and it is unfortunate that his views have not become more widely shared. In this article, I provide an argument for a corollary to Lykken's claim that the failure to achieve $p < .05$ is never a sufficient condition to reject a scientific article for publication or to conclude the finding is uninformative. Four negative consequences of refusing to publish *ns* results are discussed. Then, I outline how and when *ns* findings might be reported so as to be scientifically informative. But, this essay begins by briefly summarizing what it means for a finding to be significant or not. An example and explanation is provided for why the mere achievement of $p < .05$ might not matter much in evaluating the contribution of a finding to a scientific literature.

What $p < .05$ and *ns* Mean

A finding that is statistically significant is one that is < 5% likely to obtain if the null hypothesis of no difference or association is presumed. By convention (and not sound logic; see Levine, Weber, Hullett, Park, & Lindsey, 2008), when $p < .05$, the null hypothesis is rejected, and evidence is inferred for an alternative hypothesis (where the alternative is simply defined as "not the null").

A result that is not statistically significant at $p < .05$ is not sufficiently improbable under the statistical distribution implied by the null to satisfy the convention whereby the null is rejected. An *ns* finding is not evidence for no difference or association or a null relationship between variables. Instead, it is indeterminate with regard to the null. The statistical *nil* null hypothesis is still probably false when a finding is *ns,* but the evidence just is not strong enough to meet the conventional standard to reject the null.

Because the null is never accepted based on an *ns* finding, an *ns* finding might unfortunately be seen as uninformative. As previously mentioned, with regard to the statistical nil null hypothesis, an *ns* finding is indeterminate and unsatisfying. Neither the null, nor its alternative, can be accepted or rejected when a finding is *ns.* It follows that if the goal of the research is to provide a binary test of a statistical nil null hypotheses, and nothing more, then an *ns* finding is uninformative. It does not follow, however, that *ns* results are, therefore, meaningless more generally. Theorists, researchers, and practitioners might well be interested is assessing how some variables are related. For many (and probably most) research purposes, an up or down referendum on the null is not the point. The aim of research is often more about estimating an effect than rejecting a null. Estimating an effect provides more useful information that rejecting the statistical nil null.

Consider this hypothetical example: Researcher A finds a correlation of $r(200) = .15$, $p < .05$ (95% confidence interval [CI] = .01 to .28). Researcher B studies the same variables with an identical sample size, and finds $r(200) = .13$, $p = ns$ (95% CI = −.01 to .26). Researcher A can reject the null, but B cannot. However, both results should lead to nearly identical substantive conclusions. Further, if A's research gets published and B's does not, and this reflects a systematic bias, undesirable consequences will follow.

Publication Bias and the Systematic Distortion of Scientific Literatures

The practice of publishing only statistically significant results leads to systematic distortion of scientific evidence. Consider, again, the example in the previous paragraph. Our best guess of the population association is $\rho = .14$ (the weighted, across-study, average correlation; 95% CI = .04–.23). If $N = 202$ is a typical sample size, and many additional studies follow, then we would expect most results to fall between $r = .00$ and $r = .27$ and be quasi-normally distributed around .14. But, if only the significant results find their way into print, then the range of published values will be .15 to around .30. Keeping the smaller estimates out of the public record leads to systematic distortion. If studies vary in sample size, with many studies using smaller samples, the distortion would be more extreme due to the larger confidence limits and the increasingly upward-biased results needed for the artificial $p < .05$ standard.

This distortion stemming from the reluctance to publish *ns* results is sometimes called "publication bias." It shows up as a negative correlation between sample sizes and effects sizes in meta-analysis (Levine, Asada, & Carpenter, 2009). Because good scientific evidence is unbiased, publication bias is undesirable, and the practices that produce it should be discouraged and discontinued.

Falsifiability

A preference for supportive evidence over non-supportive evidence (confirmation bias) and for significant findings over *ns* findings also works contrary to the idea that scientific theories and hypotheses must be falsifiable. Ferguson and Heene (2012) explained this idea nicely:

> The tendency of psychological science [and communication journals] to avoid publishing null results produces a situation that limits the replicability assumption of science, as replication cannot be meaningful without the potential acknowledgment of failed replications. We argue that the field often constructs arguments to block the publication and interpretation of null results and that null results may be further extinguished through questionable researcher practices. We argue that these problems reduce psychological science's capability to have a proper mechanism for

theory falsification, thus resulting in the promulgation of numerous "undead" theories that are ideologically popular but have little basis in fact. (p. 555)

Meta-Analysis

Relatedly, failing to publish *ns* results hampers meta-analysis by providing a biased selection of primary results. The goal of accurately understanding a research topic is better obtained by triangulating results across multiple studies than by cherry-picking supportive findings. The usefulness of meta-analysis in providing an unbiased summary of findings is limited when non-supportive results are culled from the record of published research.

A major cause of *ns* findings is suboptimal statistical power. Meta-analyses maximize power. Therefore, it makes sense to allow primary findings into print so they can be cumulated with other findings to provide a stronger and more robust test than is practically possible in a primary study.

Encouraging "Questionable Research Practices"

Questionable research practices (QRPs) are things some researchers do that increase the chances for publication at the cost of the quality of the scientific inferences and that, despite the negative consequences, are not universally seen as unethical. An example might include what I call "*p*-value farming," which involves examining the outcomes of numerous significance tests, reporting only the significant or supportive tests, and reporting the results in such a way that the research consumer does not know how many tests were conducted or failed to produce $p < .05$. Like publication and confirmation biases, QRPs make results look better than they really are. The practice of favoring significant findings over *ns* results provides an incentive for engaging QRPs. Because QRPs are undesirable, the practices that motivate QRPs should be discouraged and discontinued.

Making an *ns* Finding More Meaningful

Sometimes the lack of a large difference or strong association is interesting. One instance is when an accepted theory or a widely accepted belief suggests big effects. Although an *ns* finding does not provide evidence of a zero effect, it can provide evidence that an effect is not large.

A useful way to interpret an *ns* findings is to calculate an effect size for the test and then calculate a CI around that effect size (for information on calculating effect sizes and CIs, see Levine, Weber, Park, & Hullet, 2008). The observed effect size provides an estimate of the relationship between variables, and the population effect is 95% likely to be within the CIs. Values outside the CIs are statistically improbable in the same way that the null is improbable when a result is statistically significant.

It may be tempting to argue that that an *ns* result is *ns* because of low statistical power, and that if the sample size was only larger, the findings would have been statistically significant. Although it is true, by definition, that all *ns* results lack power, presuming that a larger sample would make the finding significant is counterfactual. Effect sizes with CIs are more useful than counterfactual statistical power assertions.

Summary

Not all findings, $p < .05$ or otherwise, merit publication. However, *ns* findings can have news value, and the mere fact that a finding is statistically significant or not is *not* a key determinant of scientific contribution. Further, systematically excluding *ns* findings from the scientific record have a number of detrimental consequences that impede scientific understanding and progress. Therefore, *ns* findings should not preclude publication simply because $p > .05$.

The following is a continuation of the Lykken (1968) quote from earlier:

> The value of any research can be determined, not from the statistical results, but only by skilled, subjective evaluation of the coherence and reasonableness of the theory, the degree of experimental control employed, the sophistication of the measuring techniques, the scientific or practical importance of the phenomena studied, and so on. Editors [and readers, reviewers, authors, etc.] must be bold enough to take responsibility for deciding which studies are good and which are not, without resorting to letting the *p* value of the significance tests determine this decision. (pp. 158–159)

References

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological sciences aversion to the null. *Perspectives on Psychological Science*, *7*, 555–561.

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample size and effect size are negatively correlated in meta-analysis: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, *76*, 286–302.

Levine, T. R., Shulman, H. C., Carpenter, C. J., DeAdrea, D. C., & Blair, J. P. (2013). The impact of accusatory, non-accusatory, bait, and false evidence questioning in deception detection. *Communication Research Reports*, *30*, 169–174.

Levine, T. R., Weber, R., Hullett, C. R., Park, H. S., & Lindsey, L. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*, 171–187.

Levine, T. R., Weber, R., Park, H. S., & Hullett, C. R. (2008). A communication researchers guide to null hypothesis significance testing and alternatives. *Human Communication Research*, *34*, 188–209.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151–159.