

What though p-values conspire to cheat you: Save your greatest failure

The great failure and the greater bias

In 1887, the American scientist Albert A. Michelson and his colleague Edward W. Morley conducted the famous Michelson-Morley experiment in which they attempted to prove the existence of the luminiferous aether, through which electromagnetic waves were believed to travel. Simply stated, they failed; they found no evidence of aether's existence. However, this conclusive null result shed light on many other scientists' research, including Albert Einstein's special theory of relativity. Despite its failed result, the experiment made Michelson the first American scientist to win the Nobel Prize for Physics in 1907.¹ This leads to an important question: Is it rational to call this experiment a failure? But more importantly, what defines failure in research?

Whenever we conduct a quantitative study (usually based upon logical positivism), we instinctively look for 'statistically significant' p-values, defined as those that are less than .05 or .01. At the very moment when we confront non-significant p-values, many of us tend to fidget and declare the study a failure, although we know that the cut-off value of p (alpha level) is arbitrary at best. Some scientists simply shelve the data and move on to another study, while others change the original hypothesis or analysis approach of the study in order to achieve the desired p-value.

A potential explanation for this situation is that most research projects mandate publication as their outcome, and academic journals certainly prefer definitive results that negate the 'null hypothesis' or, at least, researchers believe this to be so. The phenomenon that only positive and statistically significant results are published is called publication bias.² This idea has proven to be fact, as research suggests that study results with low p-values (<.05) have much higher odds of being fully reported.³ For example, a recent study revealed that only 20% of null results in studies conducted between 2002 and 2012 were published, while roughly 60% of studies with strong results were published.⁴

The pathophysiology of publication bias

So, where did such a biased publication culture originate? One plausible explanation is that most researchers think that success or failure of a study is determined by the p-value. Researchers and academic journal editors alike seem to have an equation in mind: "high p-value = fails to reject the 'null hypothesis' = failure of the study." The first half of this equation may make intuitive sense, but in reality the latter is flawed. A high p-value, by definition, should only be decisive as to whether the experimenter should 'reject the null hypothesis or not,' but is not indicative of the success or failure of the study.

How else can we determine the success of a research study? Some define research as a process of scientific inquiry⁵ and, therefore, the success or failure of research is determined by whether the inquiry was thoroughly conducted and therefore adds knowledge to the world. If we agree with this definition of research, then we can clearly see that the current publication system is heavily skewed towards publishing research that produces statistically significant outcomes.

Volume 3 Issue 2 - 2016

Anselmo J Chung,¹ Heon-Jae Jeong,² Wui-Chiang Lee³

¹Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, USA

²The Care Quality Research Group, Seoul, Korea

³Department of Medical Affairs and Planning Department, Taipei Veterans General Hospital, Taiwan

Correspondence: Heon-Jae Jeong, The Care Quality Research Group, Chunjuro 174, Chuncheon, Gangwon 24450, South Korea, Tel +82-10-8878-9571, Fax +82-33-252-8558, Email hjeong1@jhu.edu

Received: January 29, 2016 | **Published:** February 02, 2016

Indeed, innumerable studies fail to reject the null hypothesis, but still contain a huge amount of information that can enrich the body of knowledge, as evident in the Michelson-Morley experiment. Thus, publication bias causes a global-level missed opportunity. As such, it is our responsibility as researchers to address this bias to promote and publish well-executed studies with rock-solid scientific methods that give meaningful information to humankind.

Breaking off the chains of publication bias

So, how can we fix this academia-wide phenomenon? First and foremost, the journals and publishers must provide a platform for researchers to safely publish non-positive results. To our knowledge, only a handful of journals (Nature Publishing Group, Journal of Cerebral Blood Flow and Metabolism, Neurobiology of Aging, and Journal of Negative Results in Biomedicine) have thus far formally addressed the problem of negative publication bias. These journals introduced a special section that provides a forum for negative results and 'promotes a discussion of unexpected, controversial, provocative and/or negative results in the context of current tenets'.^{4,6-8} Indeed, these journals are leading the way in providing full scope of knowledge in their respective scientific fields. In order for large-scale change in this field, other esteemed journals should provide similar forums for negative outcomes to be published. Journals also can provide guidelines for authors to better describe their 'seemingly unsuccessful' study results in a more convincing way, without disappointing readers.

Additionally, journals should change their article evaluation paradigm by helping to redefine failure in the scientific research world. P-values should be referenced, but they should not be the sole determinant of the success or failure of a study. Instead, p-values should indicate the direction of the proof of the study as well as the values that are meaningful in both rejecting and not rejecting the null hypothesis.

Another crucial aspect of p-value's said importance in research is peer review. If the peer review system were to change from its current rejection-prone system into a more constructive one, optimally the

best research manuscripts would be accepted regardless of whether they reject the null hypothesis or not. In this way, manuscripts that are perhaps not as traditionally impressive to readers but that exhibit solid research methods and results (regardless of p-values) can be better read and appreciated by the scientific community.

Lastly, researchers must also shift their attitude toward what constitutes the success of a study and identify creative ways to share their findings with the world, even when traditional measures of success is not met.

Failure is in the eye of the beholder: our studies may not have failed

We would like to conclude this article with a citing from one of Alexander Pushkin's famous poems (with a small modification): *What though p-values conspire to cheat you by telling that you failed, don't be dismal, don't be devastated.*⁹ The real success of any study is not defined by p-values; rather it lies beyond p-values, rather its success hinges on whether the work contributes to increasing the knowledge of the world.

Acknowledgement

None.

Conflict of interest

None.

References

1. Shankland RS. Michelson-Morley Experiment. *American Journal of Physics*. 1964;32(1):16–35.
2. Easterbrook PJ, Gopalan R, Berlin J, et al. Publication bias in clinical research. *Lancet*. 1991;337(8746):867–872.
3. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3(8):e3081.
4. Peplow M. Social sciences suffer from severe publication bias. *Nature*. 2014;15787.
5. Glanz K, Rimer BK, Viswanath K. *Health Behavior and Health Education: Theory, Research, and Practice*, 4th ed. 2008.
6. Coleman PD. Negative results can be valuable. *Neurobiol*. 2004.
7. Pfeffer C, Olsen BR. Editorial: Journal of negative results in biomedicine. *Journal of negative results in biomedicine*. 2002;1(1):2.
8. Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab*. 2006;26(12):1465–1478.
9. Chen Q. Analysis on Three Versions of If by Life You Were Deceived from Perspective of Stylistics. *Journal of Language Teaching and Research*. 2013;4(3):598–604.