ECRONICON OPEN ACCESS

EC NEUROLOGY
Mini Review

# How Stringent a P? Addressing P-Values and the Future Conduct of Translational Research in Neurology

**Owen Sisbarro[1] and James Giordano[2]***

[1]*Department of Physiology, Georgetown University Medical Center, Washington, DC, USA*

[2]*Departments of Neurology and Biochemistry, and Neuroethics Studies Program-Pellegrino Center for Clinical Bioethics, Georgetown University Medical Center, Washington, DC, USA*

***Corresponding Author:** James Giordano, Departments of Neurology and Biochemistry, and Neuroethics Studies Program-Pellegrino Center for Clinical Bioethics, Georgetown University Medical Center, Washington, DC, USA.

**Received:** May 14, 2019; **Published:** June 20, 2019

## Abstract

The groundswell of current developments in neuroscience prompts inquiry to, and address of the ways that research findings from diverse (sub)disciplines, international laboratories and clinical trials are analyzed, compared, and ultimately used. But what of the way(s) research evidence is measured and evaluated? Recently, there has been discussion - if not controversy - about the actual value of statistical evaluations conducted at the level of $p < 0.05$. Thus, we posit that the adoption and use of a new threshold (e.g. $p < 0.005$ - .001) need not invalidate prior research, but rather only indicates a change in current scientific epistemology and values of the biomedical community, and thus should occasion more detailed inquiry into the methods of previous studies, as well as the relative meaning(s) of their outcomes in light of the aims of ongoing international initiatives to develop ever more precise tools with which to assess, affect, and treat the brain and its functions and disorders.

***Keywords:*** *Research Design; Statistical Analysis; P-Value; Neurology; Neuroethics*

## Evidence and its Metrics

Much has been written about the importance of evidence in translating research to viable and valuable applications in clinical care; and evidence-based practice has become *de rigueur* in neurology [1,2]. This may be ever more the case as the groundswell of current developments in neuroscience prompts inquiry to, and address of the ways that research findings from diverse (sub)disciplines, international laboratories and clinical trials are analyzed, compared, and ultimately used [3,4]. But what of the way(s) research evidence is measured and evaluated? [5]. Recently, there has been discussion - if not controversy - about the actual value of statistical evaluations conducted at the level of $p \leq 0.05$ [6]. While p-values provide an indication of statistical significance, they do not represent nor provide a comprehensive picture of the scientific or clinical importance, relevance, or meaningfulness of the treatment effect size, and/or the research at-large. Therefore, we believe that the implication that $p \leq 0.05$ is too ambiguous, and therefore of little value, necessitates further assessment and address.

## How stringent a P?

Regarding $p \leq 0.05$ as establishing too ambiguous a threshold for statistical significance prompts inquiry about the validity, value and ultimately regard for any/all studies conducted using statistical evaluations at this level. At face value, the implications of this seem

pressing, but we posit that this may be questionable, at least to some extent. To be sure, most researchers consult a statistician, and in the majority of cases, defined procedures are used to determine which p-value can and should be employed. So the renewed call for the use of more stringent p-values, taken with extant - and perhaps newly instantiated - mathematical and statistical rules (viz. -that a p value can be set lower than ≤ 0.05 when the data is being re-directed through multiple tests) will be considered by the statistical community as a new norm [7]. A complete discussion of these mathematical and statistical literature pertinent to such protocols is beyond the scope of this essay (for overview, see Greenland., *et al.* [8]).

Still, while p-values are of value to establish and explain statistical significance, the use of a p-value should not be the most important, if not sole metric of results [6]. Here we may confront a conundrum: establishing statistical significance at $p \leq 0.05$ may (now) be seen as not sufficiently rigorous; yet the adoption of more stringent p-values (e.g. < 0.005; < 0.001) might result in important effects (and key findings and/or technological developments) being overlooked.

### Mitigating risk

Indeed, sound statistical analyses are but one element of the rigor essential to well-constructed research. Interpretation of scientific results must also be based upon the relevance and/or ramifications of the findings; effect size; and differences in end points [4,7-9]. This may be critical to studies of emerging techniques and technologies, in which sample size, study duration, and observable end points may all be relatively constrained. This is not trivial. If we are to adopt a perspective that seeks to identify benefit(s) and assess and mitigate risk(s) of new developments in brain science, then an important first step is to most accurately evaluate the capabilities, burdens and constraints incurred by a given technique or technology on a variety of levels [10].

## Conclusions

Given increasing interest in precision, and personalized medicine, and a rising tide of international initiatives in brain sciences, it may be crucial at this point in time to both engage research on emerging developments and novel use of extant technologies, and to more granularly examine how such research is conducted and its outcomes are analyzed [11,12]. In this light, we posit that the adoption and use of a new threshold (e.g. $p < 0.005$ - .001) need not invalidate prior research, but rather only indicates a change in current scientific epistemology and values of the biomedical community, and thus should occasion more detailed inquiry into the methods of previous studies, as well as the relative meaning(s) of their outcomes. Indeed, as consistent with a practical philosophy of science, it is axiomatic for the field to remain self-observant, self-critical, and self-revising in both knowledge and method(s) [13]. As so often is the case in science: new tools beget new theories, and new theories beget new tools [14,15].

The stated goals and dictates of robust international efforts in brain science (e.g. United States' Brian Research through Advancing Innovative Neurotechnologies - *BRAIN* - initiative; European Union Human Brain Project; Japan MIND/Brain Project; China Brain Project, etc.) are to develop ever more precise tools with which to assess, affect, and treat the brain and its functions and disorders [16-19]. Therefore, we believe - and argue - that there is an ethical imperative to be equally precise in the metrics and methods that are used to evaluate and validate such research findings, developments and potential applications. Thus, a more capable and adaptable statistical toolkit may be required, and it may be, as Weinberg has claimed, "time to rehabilitate the p-value" [20]. For as noted by psychologist Howard E. Gruber, in his studies of scientific creativity, "…the power and beauty of science do not rest upon infallibility, which it has not, but on corrigibility, without which it is nothing" [21]; and to quote H.G. Wolfe, sometimes "…fixity of purpose requires flexibility of method" [22].

## Acknowledgements

JG is supported by federal funds UL1TR001409 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through the Clinical and Translational Science Awards Program (CTSA), a trademark of the Department of Health and Human Services, part of the Roadmap Initiative, "Re-Engineering the Clinical Research Enterprise" The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing policies or endorsement, either expressed or

implied. Sponsors had no involvement in study design, collection, analysis, interpretation, writing and decision to and where to submit for publication consideration.

## Declarations

### Ethics Approval and Consent to Participate

Not applicable.

### Consent for Publication

Not applicable.

### Availability of Data and Material

Not applicable.

## Competing Interests

The authors declare that they have no competing interests.

## Funding

JG is supported by federal funds UL1TR001409 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through the Clinical and Translational Science Awards Program (CTSA), a trademark of the Department of Health and Human Services, part of the Roadmap Initiative, "Re-Engineering the Clinical Research Enterprise".

## Authors' Contributions

- OS drafted initial manuscript.
- JG edited and ensured relevancy of content.
- Both authors approved final submitted version.

## Bibliography

1. Demaerschalk BM and Wingerchuk DM. "Evidence-based Neurology: Management of Neurological Disorders". Oxford: Wiley-Blackwell (2015).

2. Prasad K. "Evidence-based approach in neurology practice and teaching". *Annals of Indian Academy of Neurology* 9.1 (2006): 11-19.

3. Boswell MV and Giordano J. "Evidence-based or evidence-biased? On the need to consider stakeholder values in research assessment". *Pain Physician* 12.2 (2009): 9-12.

4. Caplan LR. "Evidence based medicine: concerns of a clinical neurologist". *Journal of Neurology, Neurosurgery, and Psychiatry* 71.5 (2001): 569-574.

5. Thibault GE. "Clinical problem solving: Too old for what?" *New England Journal of Medicine* 328.7 (1993): 946-950.

6. Ioannidis JPA. "The proposal to lower p value thresholds to .005". *Journal of the American Medical Association* 319.14 (2018): 1429-1430.

7. Greenland S and Poole C. "Living with p values: resurrecting a Bayesian perspective on frequentist statistics". *Epidemiology* 24.1 (2013): 62-68.

8. Greenland S., *et al*. "Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations". *European Journal of Epidemiology* 31.4 (2016): 337-350.

9. Giordano J. "Good as gold? The randomized controlled trial: pragmatic and ethical issues in pain research". *Journal of Pain Management* 16 (2006): 68-71.

10. Giordano J. "Toward an operational neuroethical risk analysis and mitigation paradigm for emerging neuroscience and technology (neuroS/T)". *Experimental Neurology* 287.4 (2017): 492-495.

11. Giordano J. "A preparatory neuroethical approach to assessing developments in neurotechnology". *AMA Journal of Ethics* 17.1 (2015): 56-61.

12. Giordano J and Shook JR. "Minding brain science in medicine: On the need for neuroethical engagement for guidance of neuroscience in clinical contexts". *Ethics in Biology, Engineering and Medicine* 6.1-2 (2015): 37-42.

13. Suppe F. "Conclusions on the growth of scientific knowledge". In: Suppe F. (ed.) The Structure of Scientific Theories, 2nd Edition. IL: Illini Books (1977).

14. Kuhn TS. "The Structure of Scientific Revolutions". 2nd Edition. Chicago: University of Chicago Press (1970).

15. Popper KR. "Realism and the Aim of Science". London: Routledge (1983)..

16. https://www.braininitiative.org

17. https://humanbrainproject.eu

18. https://brinminds.jp/en/

19. https://en.wikipedia.org/wiki/China_Brain_Project

20. Weinberg CR. "It's time to rehabilitate the p-value". *Epidemiology* 12.3 (2001): 288-290.

21. Gruber HE. "The origin of the Origin of the Species". NY Times Book Review (1979): 7.

22. Wolfe HG. "Stress and Disease". Springfield, IL: C. Thomas (1952).

**Volume 11 Issue 7 July 2019**