

Effect size, the misnomer

Sirs,

The *p*-value is ill used. However, it is curious that in the ongoing debate and the heated action for its more proper or even no use (1-4), no consideration is given to the misnomer which, along with its confidence intervals, is the very quantity the *p*-value is related to in null hypothesis testing (NHT). The statistical jargon for this quantity is *effect size*. It is a misnomer in that it surely implies causality when used to describe study outcomes in instances where no such is indicated by the study design. On the other hand, expert statisticians, including A. Cohen who probably first used it, explicitly state that no causality is implied in an effect size (5, 6).

We propose two different schemes to improve things. The first is to keep the term "effect size" in interpreting the outcomes of all studies with a causal design in a broad sense, whether they are randomised, quasi-randomised or observational where potential confounders have been adequately addressed to the authors' and editors' satisfaction. We then discontinue the use of the term "effect size" referring to statistical analyses of non-causal studies. Note that this suggestion does not propose any different arithmetic indices from the currently used mean differences, odds ratios, correlation coefficients etc., that express to what degree the study results differ from the null in such work. It only proposes to refrain from any references to effect sizes in the texts of non-causal studies.

The second option, which we favour, is somewhat more radical. This scheme pro-

poses that we do away with "effect size" all together in all study settings. This more radical proposal has the main advantage of keeping the misnomer away from the NHT altogether and obviates any discussions and debate about whether this unfortunate jargon is suitable for the study results at hand. Either scheme, on the other hand, will remind, especially the novice, that the causality is mainly related to the study design and not the NHT.

There remains the important question what to call all those arithmetic indices that quantify how our study results differ from the null in NHT, instead of the misnomer. The terms *degree of discordance with the null* or simply *discordance with the null* are two options. Interestingly, Cohen himself almost verbatim included what we propose when he described effect size. "Without intending any necessary implication of causality, it is convenient to use the phrase effect size to mean the degree to which the phenomenon is present in the population, *or the degree to which the null hypothesis is false*." (Italics ours) (5). While we surely disagree with the first part of this definition, the wording of which is misleading in that there surely are many non-causal differences or associations in nature, we certainly agree with the wording of the second part.

Finally, as for the use of jargon in general, even when this use is not necessarily most inappropriate as in the case at hand, it has been said "We often think of it as language reserved for the brightest and most educated among us. In reality, it comes from an old French word that means 'the language and chattering of the birds'. It is unintelligible and best left in the trees." (7).

H. YAZICI¹, MD

Y. YAZICI², MD

¹Academic Hospital Istanbul, Turkey;

²New York University School of Medicine
New York, NY 10016, USA.

Please address correspondence to:

Yusuf Yazici,
New York University
School of Medicine,
333 East 38th Street,
New York, NY, 10016, USA.

E-mail: yusuf.yazici@nyulangone.org

ORCID iD:

H. Yazici: 0000-0001-6436-1639

Y. Yazici: 0000-0002-7605-5759

Competing interests: none declared.

© Copyright CLINICAL AND
EXPERIMENTAL RHEUMATOLOGY 2022.

References

1. WASSERSTEIN RL, LAZAR NA: The ASA statement on *p*-values: context, process, and purpose. *Am Stat* 2016; 70: 129-33.
2. GOODMAN SN: Aligning statistical and scientific reasoning. *Science* 2016; 352: 1180-1.
3. AMRHEIN V, GREENLAND S, MCSHANE B: Scientists rise up against statistical significance. *Nature* 2019; 567: 305-7.
4. TRAFIMOW D, MARKS M: Editorial. *Basic Appl Soc Psychol* 2015; 37: 1-2.
5. COHEN J: *Statistical Power Analysis for the Behavioral Sciences*. 2nd. Ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
6. KELLEY K, PREACHER KJ: On effect size. *Psychol Methods* 2012; 17: 137-52.
7. BANSAL S: The good, the bad and the jargon. *BRIGHT* magazine 2019 March 28; <https://bright-themag.com/the-good-the-bad-and-the-jargon-language-ngo-social-issues-journalism-b34953c9300b> (accessed July 1, 2021).