



Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en



Editorial

“Suggestive” and “Significant”: You can’t always get what you want. . .



For decades now in Medicine, and in the little world of Otorhinolaryngology, the word “significant” has been the alpha and omega, the be-all and end-all of statistical analysis in “fundamental” or just “clinical” research. This Holy Grail takes the form of a P value below a 0.05 threshold on statistical testing. Elbowing aside any notion of clinical relevance, the threshold bars the way to discussion, validates the study hypothesis and, above all, gives a green light for research to go on (and the funding to come in...).

For several years, our biostatistician friends have been forcefully arguing that a 0.05 threshold is not enough for a result to be meaningfully “significant”, and that conducting medical research on this basis creates confusion and is largely responsible for the “reproducibility crisis” that is costing society a fortune [1]. These scientists advocate setting the significance threshold at $P < 0.005$, with P values between 0.05 and 0.005 being considered “suggestive”, allowing research to continue under certain methodological conditions [1,2]. An analysis of randomized clinical studies published in 2017 in three renowned medical journals, the *JAMA*, *Lancet* and *New England Journal of Medicine*, performed by researchers in the Health Sciences Center of the University of Oklahoma, found that almost a third of the studies with results considered significant at the 0.05 threshold would be merely “suggestive”, with P between 0.05 and 0.005, if this system were adopted by the medical community [2].

This shoddy use of statistics in medical articles, which some would not hesitate to call misuse [3], is compounded by another phenomenon that was recently exposed: pressure exerted by researchers on the biostatisticians who work with (or is it “for”?) them, as revealed by a study based on responses to a questionnaire sent to more than 500 members of the American Statistical Association [4]. The authors highlighted some 20 abuses in requests made by researchers to biostatisticians [4]. The four most common devious requests were: deleting or altering data so as to back up the study hypothesis, interpreting the statistical tests according to expectations rather than the actual results, omitting key missing data that might bias the results, and failing to adhere to the initial hypothesis [4]. The study did not say how the biostatisticians responded to these “inappropriate” requests; but it certainly showed how much room for improvement there is in medical researchers’ training in methodology and ethics, going well beyond the simple understanding of the “significant” or “suggestive” status of a P value. This is all the more important as bad research practices set in very early in a scientific career. In Germany, an article analyzing the prevalence of dubious research practices and their predictive factors, in a population of more than 200 students and recent graduates in psychology, stressed the role of the thesis supervisor in developing their students’ critical faculties regarding

Box 1: The 20 ‘inappropriate’ requests made by North American researchers to their biostatisticians (from [4]), by decreasing order of severity in the biostatisticians’ opinion.

- Falsify the statistical significance (such as the P value) to support a desired result.
- Change data to achieve the desired outcome.
- Remove or alter some data records (observations) to better support the research hypothesis.
- Interpret the statistical findings on the basis of expectations, not the actual results.
- Do not fully describe the treatment under study because protocol was not exactly followed.
- Ignore violations of assumptions because the results may change to negative.
- Modify a measurement scale to achieve some desired results, rather than adhering to the original scale as validated.
- Do not report the presence of key missing data that could bias the results.
- Report power on the basis of a post hoc calculation, but make it seem like an a priori statement.
- Request to not properly adjust for multiple testing when “a priori, originally planned secondary outcomes” are shifted to an “a posteriori primary outcome status”.
- Conduct too many post hoc tests, but purposefully do not adjust alpha levels to make results look more impressive than they really are.
- Remove categories of a variable to report more favorable results.
- Do not mention interim analyses to avoid “too much testing”.
- Report results before data have been cleaned and validated.
- Do not discuss the duration of follow-up because it was inconsistent.
- Stress only the significant findings, but underreport non-significant ones.
- Do not report the model statistics (including effect size in Anova or R^2 in linear regression).
- Do not show the plot because it did not show an effect as strong as you had hoped.

malpractice and the will to perform quality research [5]. It was not the students’ belief in the need to have “significant” results in order to have a “good” study and/or get a good grade that correlated with their perception or practices regarding “dubious” research; rather, it was the feeling they had about their supervisor’s thoughts on the matter that affected their behavior (Box 1).

Never forget: *Errare Humanum Est, Perseverare Diabolicum!*

Disclosure of interest

The authors declare that they have no competing interest.

References

- [1] Laccourreye O, Maisonneuve H. Toward a new threshold for the *P* value? *Eur Ann Otorhinolaryngol Head Neck Dis* 2018;135:299.
- [2] Wayant C, Scott J, Vassar M. Evaluation of lowering the *P* value threshold for statistical significance from .05 to .005 in previously published randomized clinical trials in major medical journals. *JAMA* 2018;320:1813–4.
- [3] Localio AR, Stack CB, Meibohm AR, et al. Inappropriate statistical analysis and reporting in medical research: perverse incentives and institutional solutions. *Ann Intern Med* 2018;169:577–8.
- [4] Wang MQ, Yan AF, Katz RV. Researcher requests for inappropriate analysis and reporting: a U.S. survey of consulting biostatisticians. *Ann Intern Med* 2018;169:554–8.

- [5] Krishna A, Peter SM. Questionable research practices in student final theses—Prevalence, attitudes, and the role of the supervisor's perceived attitudes. *PLoS One* 2018;13(8):e0203470, <http://dx.doi.org/10.1371/journal.pone.0203470> [eCollection 2018].

O. Laccourreye^{a,*}

^a HEGP, service d'oto-rhino-laryngologie et de chirurgie cervico-faciale, université Paris Descartes Sorbonne Paris Cité, AP-HP, 20-40 Rue Leblanc, 75015 Paris, France

H. Maisonneuve^b

^b 30, rue Faidherbe, 75011, Paris, France

* Corresponding author.

E-mail address: ollivier.laccourreye@aphp.fr
(O. Laccourreye)