

**Letter to Editor****“A statistically non-significant difference”: Do we have to change the rules or our way of thinking?****Kafi, M.<sup>1\*</sup> and Ansari-Lari, M.<sup>2</sup>**<sup>1</sup>Department of Clinical Sciences, School of Veterinary Medicine, Shiraz University, Shiraz, Iran; <sup>2</sup>Department of Food Hygiene and Public Health, School of Veterinary Medicine, Shiraz University, Shiraz, Iran**\*Correspondence:** M. Kafi, Department of Clinical Sciences, School of Veterinary Medicine, Shiraz University, Shiraz, Iran. E-mail: kafi@shirazu.ac.ir

10.22099/IJVR.2022.44044.6470

(Received 15 Jun 2022; accepted 22 Aug 2022)

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)**Dear Editor,**

Recent years have witnessed a lot of debate about statistical significance and the use of p-value in research. Some strongly believe that p-value and statistical significance should be left out (Amrhein *et al.*, 2019), while others suggest more stringent thresholds for statistical significance (Benjamin *et al.*, 2018; Ioannidis, 2018); both are two extremes of a spectrum, we believe. On the one hand, there is a concern about the poor reproducibility of the published results (false positive findings) which have sometimes been ascribed to the widespread use of point-and-click data-analysis software. This approach has made it easy for researchers to sift through massive data sets without fully understanding the analyzing method resulting in small p-values that may not actually mean anything (Nuzzo, 2015; Ioannidis, 2018). On the other hand, there is the publication bias or the so-called file drawer problem. The nature of the bias is that studies that do not produce a statistically significant result are less likely to be published than those that do produce a statistically significant result, even if the results are not biologically or economically significant. In this line, Amrhein *et al.* (2019) published a letter to the editor in Nature, calling for a stop to the use of p-values as a sole criterion to decide whether a result refutes or supports a scientific hypothesis. This letter to the editor was supported by 854 signatories. Interestingly, more than 58% (494/854) of the signatures came from public health scientists. And more importantly, more than 17% (142/854) of the signatories were experts in mathematics, biostatistics, and epidemiology.

This topic requires further investigation and discussion. For example, decades after decades, many reproduction researchers around the world are evaluating the effects of different hormonal treatments and protocols on the pregnancy rate and reproductive performance in different species of animals, mainly

ruminants. The pregnancy rate is a variable most often limited between 30 and 45% in such studies in the bovine. We use Chi-square test to statistically analyze the proportions of pregnancies, and in some instances, we obtain p-values more than 0.05. Then we respectfully obey the mathematical output and politely say there is no statistically significant difference between the groups, while the difference in the pregnancy rate between the groups was more than 10%, for example. This mathematical expression yields no benefit to the dairy industry, reduces the chances of publication, and ultimately burns and buries the researcher's novel hypothesis. Bovine practitioners all know how profitable a 10% increase in the pregnancy rate is for dairy industry assuming to manage a dairy herd of 1000 milking cows. Increasing the 21-day pregnancy rate by 1% could increase the net profit return 14 US\$/cow per year. The average value for each pregnancy has been estimated as \$278, increasing to more than \$400 between the fifth and eighth months after calving in high producing dairy cows in the USA (De Vries, 2004; Cabrera, 2014). Who or what should be really blamed when we face with an embarrassing “statistically non-significant difference” output? The nature with its inherent variabilities, the rule makers, the rules or ourselves? The answer is, undoubtedly, “us” that do not consider the rules from the beginning of the research. A “statistically non-significant difference” is significant and valuable, and no one deserves to be criticized as long as the rules have sufficiently been attended during all stages of the research, we believe. To scientifically convince ourselves, editors and the readers when we obtain a “statistically non-significant difference”, the following points and guidelines are proposed:

1- A statistically significant relationship or effect does not mean or imply that a relationship or effect is highly probable, real, true, biologically or economically significant. Some statistically analyzed findings are highly significant but are not clinically significant. Also,

a statistically non-significant relationship or effect does not mean or imply that a relationship or effect is really absent, false, or unimportant. A non-significant result can be clinically or even economically significant even with insufficient statistical power of the study (Wasserstein and Lazar, 2016; Wasserstein *et al.*, 2019). In this context, Amrhein *et al.* (2019) reported that 51% (402/791) of the published papers in five prestigious journals erroneously interpreted “non-significant statistical findings” as indicating “no effect” (Supplementary information to: Retire statistical significance (Comment in Nature 567, 305-307; 2019).

2- We should realize that p-value is not merely a figure, but a continuum; in other words, it is not a binary expression showing either significance or non-significance. Therefore, it is better to focus on the magnitude of the effect not sticking solely to p-value and interpret it as whether an effect exists or not (Ioannidis, 2019). The results either statistically significant or non-significant should be clearly interpreted emphasizing uncertainty in them via reporting the confidence intervals (Amrhein *et al.*, 2019).

3- Designing a study is a multi-stage process. The first and the most vital stage in designing a research is the power of the study using an appropriate sample size. A common pitfall to avoid is designing and carrying out a research work solely based on the similar published papers with statistically significant results even with small sample sizes. The simplicity of the researcher is here; mimicking what has been performed in those papers. They may not realize the possible removing effects of other confounding factors including the strain, age and sex of animals by using a uniform number of animals in those studies. Other possibilities in these published studies with low power include papers published based on relations with the editor-in-chief or the editorial board or due to the presence of a well-known researcher as a co-author. We have to deeply understand that the rules of statistical analysis have been objectively set up by statisticians during last decades while the shortcoming trait of “subjectivity” has not been eliminated when research papers are evaluated by journal referees and even by peer reviewers. P-value does not recognize how important the findings of research are, but a reviewer can report the importance of the findings whether statistically significant or not. P-value objectively shows the significant or non-significant difference among experimental groups of a study, but how can we ignore the reviewer’s subjectivity in evaluating the manuscript? The subjectivity of the paper evaluation is evident by the fact that a journal publishes a paper which has been rejected by another journal while both journals have the same scope of publication, similar quality ranking, and peer-reviewing process.

4- The above arguments also emphasize the research community is in dire need of training about statistical abuse and misconception. The training should start with ourselves. We all have to accept the holy rules of complexities in the experimentation rather than looking at how others are getting something published. In

addition, the scientific community including researchers and particularly referees and journal editors have responsibility for ongoing monitoring of the literature for the right application of the statistical methods and the way the results, either significant or non-significant, are reported and interpreted (Leek *et al.*, 2017; Amrhein *et al.*, 2019).

To sum up, the low risk rule is to stick to the rules rather than solely doing research based on what has already been published even in prestigious journals. We advise ourselves to practice to replace statistical significance with statistical thinking (Wasserstein *et al.*, 2019). This, indubitably, means conducting studies using well-defined rules while trying to present and interpret the finding soundly considering the biological and economic significance, as well. As such, the real position of p-value and the applicable meaning of our findings are all acknowledged by readers, health policy makers as well as agriculture economists.

## Acknowledgement

This research was funded by Shiraz University.

## Conflict of interest

We declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

## References

- Amrhein, V; Greenland, S and McShane, B (2019). Retire statistical significance. *Nature*. 567: 305-307. <https://doi.org/10.1038/d41586-019-00857-9>.
- Benjamin, DJ; Berger, JO; Johannesson, M; Nosek, BA; Wagenmakers, EJ; Berk, R and Johnson, VE (2018). Redefine statistical significance. *Nat. Hum. Behav.*, 2: 6-10. <https://doi.org/10.1038/s41562-017-0189-z>.
- Cabrera, VE (2014). Economics of fertility in high-yielding dairy cows on confined TMR systems. *Animal*. 8: 211-221.
- De Vries, A (2004). Economics of delayed replacement when cow performance is seasonal. *J. Dairy Sci.*, 87: 2947-2958.
- Ioannidis, JPA (2018). The proposal to lower P value thresholds to .005. *JAMA*. 319: 1429-1430. <https://doi.org/10.1001/jama.2018.1536>.
- Ioannidis, JPA (2019). The importance of predefined rules and prespecified statistical analyses: do not abandon significance. *JAMA*. 321: 2067-2068. <https://doi.org/10.1001/jama.2019.4582>.
- Leek, J (2017). Five ways to fix statistics. *Nature*. 551: 557-559. <https://doi.org/10.1038/d41586-017-07522z>.
- Nuzzo, R (2015). Fooling ourselves. *Nature*. 526: 182-185. <https://doi.org/10.1038/526182a>.
- Wasserstein, RL and Lazar, NA (2016). The ASA’s statement on p-values: context, process, and purpose. *Am. Stat.*, 70: 129-133. <https://doi.org/10.1080/00031305.2016.1154108>.
- Wasserstein, RL; Schirm, AL and Lazar, NA (2019). Moving to a world beyond “ $p < 0.05$ ”. *Am. Stat.*, 73: 1-19. <https://doi.org/10.1080/00031305.2019.1583913>.