Organisms
Journal of Biological Sciences

COMMENTARIES

Vol. 3, 1 (July 2019)
ISSN: 2532-5876
DOI: 10.13133/2532-5876_5.5

Put the blame on the formula: an incredible (but real) tale from the top of modern science

Alessandro Giuliani [*]

[*] *Istituto Superiore di Sanità, Roma, Italy*

Corresponding author: Alessandro Giuliani, alessandro.giuliani @iss.it

Citation: Giuliani, A, 2019, "Put the blame on the formula: an incredible (but real) tale from the top of modern science", *Organisms. Journal of Biological Sciences*, vol. 3, no. 1, pp. 17-19. DOI: 10.13133/2532-5876_5.5

Commentary on
Amrhein, V, Greenland, S, McShane, B, 2019, "Scientists rise up against statistical significance", *Nature*, 567, 305-307 (2019) doi: 10.1038/d41586-019-00857-9

An old tale

It is common, in critical situations like famine or epidemics, to look for someone to blame so to relief the sense of frustration coming from an unexplainable tragedy.

In Italian literature, we can read a very clear, lucid and up to date account of this situation, in the '*Storia della Colonna Infame*' by Alessandro Manzoni (Manzoni A., 1843). The readers not familiar with Italian language can read the '*History of Infamy Column*' here: https://archive.org/stream/betrothedlovers00manzgoog#page/n907/mode/2up).

The tale deals with a real event happened during the terrible 1630 plague in Milan: two innocent people were death sentenced as 'plague spreaders' (*untori* in Italian) so to find someone to blame of the terrible epidemics and keep the population calm preventing possible riots. The actual science plague is the widely recognized reproducibility crisis of biomedicine, especially cruel in discovery oriented research and observational studies (see Ioannidis, 2005; Young and Karr 2011; Young 2017).

As aptly pointed out by Nuzzo (Nuzzo 2014) and by Young and Karr (Young and Karr 2011), the statistical fallacy provoking (together with other non-statistical causes linked to biological theory that here should be out-of-scope to enumerate (Noble 2011) the crisis is linked to a gross misunderstanding of the meaning of statistical significance by biomedical scientists. The reaching of a $p<0.05$ significance is an instrument (namely an empirical formula) designed in the thirties for assigning a 'worthy of second look' label to an experimental result (Nuzzo 2014) but was elevated to the rank of 'definitive truth' by biomedical sciences of the last decades. Scientists, when in presence of the magic number behave as a 'Deer in Headlights' (Young and Karr 2011). A deer caught in the headlights freezes, much like an author or reader seeing a p-value < 0.05, and think there must be a real effect.

This absurd and irrational behaviour prompted the American Statistical Society to produce a document (Wasserstein and Lazar 2016) restating 80 years old basic statistics concepts to provide six 'Principles to Improve the Conduct and Interpretation of Quantitative Science'. The problem is that the authors of the commented paper are not satisfied by the Wasserstein and Lazar principles, they quote this paper but they want to

go ahead and '*call for the entire concept of statistical significance to be abandoned*'. They seem to ignore the huge literature pointing to the separation between clinical and statistical significance (Kraemer et al. 2003). The dependency (the width of confidence interval goes to zero at increasing sample size (*n*), for the simple reason it has the square root of *n* at the denominator) of p-value from sample size makes possible to reach in any case the statistical significance for *n* going to infinite. This is not a failure in the formula; it is perfectly true that Joe, whose height is 183.15 cm, is taller than Mike who is 183.09 but a 0.6 mm difference is irrelevant. The same happens for a drug decreasing in all the patients the blood pressure of 0.1 mm of Hg. We will get a striking statistical significance in a pre- / -post paired t-test (even keeping *n* relatively small) and THIS MUST BE THE CASE as for statistics, but the effect size is negligible from the clinical standpoint and our drug is ineffective.

The mysterious p-value

P-value indicates the probability that an outcome as extreme as the observed one could happen, if the null hypothesis (no effect) were true. Statistical significance, however, does not provide information about the strength of the relationship (effect size) or whether the relationship is meaningful (clinical significance) (Kraemer et al. 2003). This is why, before the experimentation, is crucial to perform a power analysis to decide the needed sample size in order to reach a statistical significance, in the case the effect is greater than a previously declared *minimal* effect size assuring the clinical relevance.

This is not rocket science (all these notions are discussed in any introductory statistics course) but the authors of the commented paper seem to prefer another way. They claim that if the effect they look for is '*within the confidence interval*' (as appreciated by eye), this is enough to define it as 'relevant', totally discarding the possibility it could be there by chance.

I think John Ioannidis (Ioannidis, 2019) perfectly stated the point in his correspondence to the Amrhein's paper (Amrhein *et. al,* 2019). The Ioannidis correspondence title perfectly summarizes the effect on science of the Amrheim *et al* proposal '*Retiring statistical significance would give bias a free pass*'.

John Ioannidis is not a 'reactionary defendant' of the old order, on the contrary, he was the first to raise the problem of irreproducibility crisis in 2005 (Ioannidis, 2005), but his commentary is drastic, *'Statistical significance sets a convenient obstacle to unfounded claims. In my view, removing the obstacle (V. Amrhein et al. Nature 567, 305–307; 2019) could promote bias. Irrefutable nonsense would rule'.*

Ioannidis last remark is again crucial and I would like to start from there (having nothing else to add as for the statistical side) '*If rules are set before data collection and analysis, then statistical guidance that is based on appropriate thresholds is helpful. However, post hoc and subjective statistical inference is susceptible to conflicts of interest. A company could, for example, claim that any results somehow support licensing of its product'*.

Conundrums

Well, here we have a proposal ignoring basic statistical principles, signed by 800 scientists and appearing on Nature, claiming for a dismissing of any statistical significance. On the other side, we have a biotechnological and pharmaceutical industry experiencing a big crisis in terms of declining of efficacy of research. On the regulatory science side, we experience the concomitant political need of 'evidence based' limits to the approval of potentially toxic chemicals and/ or industrial practices in presence of an evident crisis of descriptive epidemiology (Young, 2017). In addition, we could mention the regulatory problems raised by unconventional potentially toxic agents like nanoparticles asking for a very new testing approach needing a still to come theory of physical action on biological systems (Nigro et al. 2018).

Now the reader could try and formulate his/her hypotheses on the reason(s) pushing 800 top scientists to behave in a way similar to the uneducated people of Milan during the 1630 plague, asking the death sentence for a formula that (like any formula) must be used with consciousness and not like the Delphi oracle.

In order to avoid any misunderstanding (here capital letters are mandatory, this is disclaimer) I **DO NOT** THINK THE AUTHORS (and the scientists that signed the proposal) ARE IN BAD FAITH. On the contrary **I BELIEVE THEY** ARE PUSHED BY THE NEED OF SCIENTIFIC ADVANCEMENT but, given the above-sketched context of efficacy crisis and emergent problems, their proposal could produce deleterious societal effects.

It is worth noting that the two innocent people death sentenced after terrible tortures in 1630, Guglielmo Piazza and Giacomo Mora, were a public health civil servant and a barber (the equivalent of nowadays sur-

geon). That is to say, two members of the 'science and regulatory' community: trying not to overtly face the actual science crisis while relying on tricks could, on the long run, give rise to a widespread refuse of scientific culture.

This is a very dangerous path.

Conflict of Interest

The Author has no proprietary, financial, professional or other personal interest of any nature in any product, service or company. The Author alone is responsible for the content and writing of the paper.

References

Amrhein V, Greenland S, McShane B. (2019) Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307. doi: 10.1038/d41586-019-00857-9.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124. doi.org/10.1371/journal.pmed.0020124.

Ioannidis, J.P. (2019) Retiring statistical significance would give bias a free pass. *Nature*. 567 (7749), 461. doi: 10.1038/d41586-019-00969-2.

Young, S. S., & Karr, A. (2011). Deming, data and observational studies: a process out of control and needing fixing. *Significance*, *8*(3), 116-120. doi.org/10.1111/j.1740-9713.2011.00506.x

Young, S. S. (2017). Air quality environmental epidemiology studies are unreliable. *Regulatory Toxicology and Pharmacology*, *86*, 177-180. doi.org/10.1016/j.yrtph.2017.03.009

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(12), 1524-1529. DOI: 10.1097/01.chi.0000091507.46853.d1.

Manzoni A. (1843) *Storia della Colonna Infame*, Edizioni ETS Milano, 2009 EAN: 9788846723307

Nigro, A., et al. (2018) Dealing with Skin and Blood-Brain Barriers: The Unconventional Challenges of Mesoporous Silica Nanoparticles. *Pharmaceutics 10.4* 250. doi.org/10.3390/pharmaceutics10040250.

Noble, D. (2011). A theory of biological relativity: no privileged level of causation. *Interface focus*, *2*(1), 55-64. https://doi.org/10.1098/rsfs.2011.0067

Nuzzo, R. (2014). Scientific method: statistical errors. *Nature News*, *506*(7487), 150. doi:10.1038/506150a.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133. doi.org/10.1080/00031305.2016.1154108.