STUDIA DOCTORALIA
PSYCHOLOGY AND EDUCATIONAL SCIENCE
UNIVERSITY OF BUCHAREST
VIRTUTE ET SAPIENTIA

EDITORIAL

p-Hacking as a Questionable Research Practice in Industrial and Organizational Psychology

Bogdan Cocoș

https://doi.org/10.47040/sd0000084

In light of the pandemic context, more and more people, both specialists and researchers, as well as people interested in accurate and up-to-date information, turn to the scientific literature for guidance, recommendations and predictions. This fact is desirable, not through the authority of science as a body of knowledge, but through its understanding as a method (i.e., of the many) as objective as possible of understanding reality. However, we may be entitled to question the method itself, especially at times when evidence may become non-evidence (e.g., the publication in top academic journals of questionable articles in terms of research practices, followed by retractions; see Sengupta, 2020).

In the scientific literature, such a phenomenon is called questionable research practices (i.e., or QRPs), which can be understood in terms of unethical methods of searching for evidence, in order to fraudulently support certain hypotheses (John, Loewenstein, & Prelec, 2012).

Of course, when such a concept is isolated from the reality in which it exists, the first curiosity that may arise is that of prevalence rates. For example, Banks, Rogelberg, Woznyj, Landis, and Rupp (2016), based on 64 articles in psychology and related fields, estimated a QRP prevalence rate of 91%. Additionally, 88% of Italian psychologists have reported at least one QRP-type deviation during their academic career, consistent with data collected in the United States and Germany (Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017). A little more targeted, Crede and Harms (2019) estimate prevalence rates of over 90% of QRPs in the field of organizational research.

All this evidence converges not to discredit the scientific research process, but to optimize it. Thus, as Kepes and McDaniel (2013) note, psychology still faces obstacles to the integrity and credibility of the scientific process; at that time, various reactions may arise regarding the credibility of the scientific literature in the industrial and organizational field.

p-Hacking

Definition

One of the possible reactions regarding the authenticity of the scientific research process in the industrial and organizational field refers to the concept of p-hacking. It can be operationally defined as a practice error in which researchers collect or use the collected data until a statistically insignificant result becomes significant (Head, Holman, Lanfear, Khan, & Jennions, 2015).

p-Values and Statistical Significance

This pursuit of statistical significance is a disappointing reality of the contemporary academic context. More and more students and junior researchers are being trained to actively seek “statistically significant” p-values, despite the fact that this type of significance is not a satisfactory

metric of the real significance of some results. Additionally, it is known that p-values do not represent an effect size index or a predictive index of the degree of importance of an obtained result (Wasserstein & Lazar, 2016). However, it is hypothesized that the elimination of such a concept, in the context of academic prestige (i.e., quantified by the number of publications), is unlikely (Head et al., 2015).

From the arguments stipulated so far, it is easy to understand that a higher incidence rate of p-hacking has the potential to generate more false-positive scientific literature (Brodeur, Cook, & Heyes, 2020). In the Null Hypothesis Statistical Testing (i.e., NHST) framework, the false-positive error refers to the Type I statistical error (i.e., also called alpha), by which a true null hypothesis is wrongly rejected (Kim, 2015). As an example in industrial and organizational psychology, a researcher could look for the incremental validity of openness to experiences over general mental ability, in the context of staff selection. By repeating the same hierarchical regression analysis 100 times (i.e., and changing certain aspects, therefore manipulating some of the analyses), according to NHST, the researcher could obtain 5 cases (i.e., at a unilateral or bilateral alpha threshold of .05) in which his hypothesis was erroneously confirmed. The erroneous conclusion will also be that openness to experiences has statistically significant incremental validity over general mental ability, in the context of staff selection. Of course, this example is a handy one, being, realistically, nothing more than an analogy (i.e., bearing a high degree of reductionism) for the practice of everyday research in the field of industrial and organizational psychology. Therefore, the discussion about p-hacking can guide us in the direction of the incompatibility between the data collected and analyzed by a researcher and reality. Often, this incompatibility is attributed to the concept of chance or luck, despite the fact that Wasserstein and Lazar (2016) argue against such an operational definition of the p-value.

Associated Issues

Thus, some of the consequences of p-hacking may be the undermining of the robustness of scientific processes, the depletion of temporal (i.e., after questionable evidence has been implemented in practice, it will take a considerable amount of time for a recovery) or financial (i.e., investing money in practices or policies supported by questionable evidence) resources, and, of course, the amplifying skepticism or lack of credibility, when it comes to science, as a systematic body of knowledge.

The scientific literature recommends, in this regard, an adjustment of the research process, towards the best currently available practices regarding transparency and accuracy applied in the field of industrial and organizational psychology (e.g., Efendic & van Zyl, 2019; Woo, O'Boyle, & Spector, 2017).

Discussion

In the light of the above arguments, the scientific literature recommends, in addition, not the attempt to eliminate p-hacking, but an optimized proactive training of researchers, with guidelines to clarify acceptable practices (e.g., preregistration of studies, measurement of only important variables for research, the use of appropriate sample sizes, an emphasis on the methodological and statistical quality of the study, to the detriment of statistical significance or novelty) and questionable scientific research (Head et al., 2015).

In Romania, there does not necessarily seem to be the needed or satisfactory ground for discussing such issues related to the integrity of scientific research. Despite this, this article will propose a series of online resources of maximum interest for educating research practices in optimal directions from an ethical standpoint.

Next, important resources will be listed for verifying the accuracy of statistical reporting (http://statcheck.io/; Rife, Nuijten, & Epskamp, 2016) and for the applied understanding and simulation of the p-hacking concept (https://shinyapps.org/apps/p-hacker/; https://shinyapps.org/apps/p-checker/; Schönbrodt, 2016, 2018). These can be useful for both personal use, and future researchers' education.

Conclusion

Thus, taking into account all the aforementioned aspects, the research practice in industrial and organizational psychology could benefit from training researchers on questionable practices, such as p-hacking. It is worth mentioning that the existence of resources is not problematic, but the low motivation to use them and to understand the processes and mechanisms underlying them is. At the same time, scientific literature can no longer be seen as an authority, but as a collective effort to optimize knowledge, especially in a field that informs and guides specific and complex public processes, such as industrial and organizational psychology.

REFERENCES

Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PloS One, 12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology, 31*, 323-338. https://doi.org/10.1007/s10869-016-9456-7

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review, 110*(11), 3634-60. https://doi.org/10.1257/aer.20190687

Crede, M., & Harms, P. (2019). Questionable research practices when using confirmatory factor analysis. *Journal of Managerial Psychology, 34*(1), 18-30. https://doi.org/10.1108/JMP-06-2018-0272

Efendic, E., & van Zyl, L. E. (2019). On reproducibility and replicability: Arguing for open science practices and methodological improvements at the South African Journal of Industrial Psychology. *SA Journal of Industrial Psychology, 45*(1), 1-10. http://doi.org/10.4102/sajip.v45i0.1607

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol, 13*(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524-532. https://doi.org/10.1177%2F0956797611430953

Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in industrial and organizational psychology?. *Industrial and Organizational Psychology, 6*(3), 252-268. https://doi.org/10.1111/iops.12045

Kim, H. Y. (2015). Statistical notes for clinical researchers: Type I and type II errors in statistical decision. *Restorative Dentistry & Endodontics, 40*(3), 249-252. https://doi.org/10.5395%2Frde.2015.40.3.249

Rife, S. C., Nuijten, M. B., Epskamp, S. (2016). statcheck: *Extract statistics from articles and recompute p-values* [web application]. Retrieved from http://statcheck.io

Schönbrodt, F. D. (2016). *p-hacker: Train your p-hacking skills!* [web application]. Retrieved from http://shinyapps.org/apps/p-hacker/

Schönbrodt, F. D. (2018). *p-checker: One-for-all p-value analyzer* [web application]. Retrieved from http://shinyapps.org/apps/p-checker/

Sengupta, S. (2020). An editorial perspective on the infamous COVID-19 studies retracted by Lancet and NEJM. *Indian Journal of Ophthalmology, 68*(7), 1247-1248. https://doi.org/10.4103%2Fijo.IJO_1853_20

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

Woo, S. E., O'Boyle, E. H., & Spector, P. E. (2017). Best practices in developing, conducting, and evaluating inductive research. *Human Resources Management Review, 27*(2), 255-264. https://doi.org/10.1016/j.hrmr.2016.08.004

Author Note

The following manuscript was translated from a Romanian term paper for one of the author's master's programme classes, Academic Ethics & Integrity (Etică și integritate academică).