

A method to streamline p -hacking

Ian Hussey

The analytic strategy of p -hacking has rapidly accelerated the achievement of psychological scientists' goals (e.g., publications & tenure), but has suffered a number of setbacks in recent years. In order to remediate this, this article presents a statistical inference measure that can greatly accelerate and streamline the p -hacking process: generating random numbers that are $< .05$. I refer to this approach as *pointless*. Results of a simulation study are presented and an R script is provided for others to use. In the absence of systemic changes to modal p -hacking practices within psychological science (e.g., worrying trends such as preregistration and replication), I argue that vast amounts of time and research funding could be saved through the widespread adoption of this innovative approach.

p -hacking – the updating or adjusting data or analyses in light of prior beliefs about hypotheses – has proven to be of exceptional utility to the goals of psychological scientists (e.g., acquiring high-impact publications, tenure, and paid speaking engagements). While a number of useful tutorials in p -hacking and related strategies exist (e.g., Bakker et al., 2012; Simmons et al., 2011), insightful commentators have pointed out that only those with a ‘flair’ for it are likely to make it in the world of psychological science (Baumeister, 2016). However, progress has slowed in recent years due to a number of unfortunate setbacks, including wider use of replication and pre-registration (e.g., Munafò et al., 2017; Open Science Collaboration, 2015) by methodological terrorists (Fiske, 2016) and data parasites (Longo & Drazen, 2016).

In this article, I introduce the *pointless* metric and demonstrate how it can streamline the process of p -hacking your results. While this metric does suffer from the mild flaw of providing zero diagnosticity of the presence or absence of a true effect, this property is largely irrelevant to most psychological scientist's primary goals (i.e., publishability: Nosek et al., 2012). Secondary goals such as valid and useful insights into human behaviour are also occasionally met, albeit incidentally. More importantly, the metric possesses three superior characteristics. First, it is non-inferior to current p -hacking practices, which also tell us little about the presence or absence of a true effect (large scale replications put this diagnosticity at no better than a coin toss: Klein et al., 2018; Open Science Collaboration, 2015). Second, it retains a far more

important property of hacked p values: by guaranteeing significant results, it maintains predictive validity for publishability. Finally, it also provides economic benefits relative to the high total life-cycle costs associated with traditional p -hacking (e.g., by eliminating the need for comprehensive graduate training in either statistics or ‘flair’ for p -hacking).

Methods and results

I observed that traditional approaches are relatively time consuming and inefficient (i.e., exploitation of researcher degrees of freedom until $p < .05$: Simmons et al., 2011). The *pointless* metric was inspired by the observation that, regardless of the specific p -hacking strategy employed, the product of this process is highlight reliable (i.e., the statistical result “ $p < .05$ ”). As such, many intermediary steps are therefore arguably unnecessary, and the same end result can be obtained more efficiently by automation. This is accomplished by generating a random number that is $< .05$. I recommend researchers to refer to this this statistical inference procedure as a form of machine learning to increase their chances of getting published. R code to calculate *pointless* is provided below:

```
p_pointless <- runif(1, 0, 0.0499)
print(paste("p_ointless =", p_ointless))
```

To evaluate the performance of this highly advanced machine learning procedure compare hacked p values, I performed a simulation study. In line with

modal p -hacking practices, only the key property of diagnosticity for publishability (i.e., $p < .05$) was considered. 10,000 cases were simulated (see Appendix for R code). Results demonstrated the results of *pointless* and traditional p -hacked results are congruent in 100% of cases. Although variation in individual coefficients frequently differ by large margins, both strategies satisfy the core criterion of producing significant results. More importantly, execution time for *pointless* is less than one second, whereas traditional p -hacking techniques can take hours or days to apply – not to mention years of training in the normalization of p -hacking practices.

Discussion

Traditional p -hacking involves starting with a sound analytic strategy and then iteratively degrading this until the results support one's hypothesis. On the basis that this strategy almost invariably returns significant results, many burdensome aspects of this analytic process can arguably be bypassed via automation. The most parsimonious method was selected: random number generation. Results from a simulation study demonstrate that decision making on the basis of traditional hacked p values and *pointless* are equivalent, and that the latter requires several orders of magnitude less time and resources to calculate. Academic productivity can therefore be greatly increased through the widespread adoption of this approach.

Now that the data processing and analytic process has been streamlined, future work should consider whether data collection itself may be an inefficient use of researchers' time or even redundant. A pilot study by Prof Diederik Stapel suggests that primary goals (e.g., tenure) can indeed be achieved without it (Verfaellie & McGwin, 2011).

References

- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543–554. <https://doi.org/10.1177/1745691612459060>
- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153–158. <https://doi.org/10.1016/j.jesp.2016.02.003>
- Fiske, S. T. (2016). Mob Rule or Wisdom of Crowds? *APS Observer, Advance online draft*. http://datacolada.org/wp-content/uploads/2016/09/Fiske-presidential-guest-column_APS-Observer_copy-edited.pdf
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Longo, D. L., & Drazen, J. M. (2016). Data Sharing. *New England Journal of Medicine*, 374(3), 276–277. <https://doi.org/10.1056/NEJMe1516564>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021-1>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Verfaellie, M., & McGwin, J. (2011). The case of Diederik Stapel. *APA Psychological Science Agenda*. <https://www.apa.org/science/about/psa/2011/12/diederik-stapel>

Appendix: R code for simulation

```
simulation <- function() {
  # pointless
  p_pointless <- runif(1, 0, 0.0499)
  if(p_pointless < 0.05) {
    publishable_p_pointless = TRUE
  } else {
    publishable_p_pointless = FALSE
  }

  # traditional (hacked) p values
  # set to upper bound of observable
  p <- 0.049
  if(p < 0.05) {
    publishable_p = TRUE
  } else {
    publishable_p = FALSE
  }

  # compare
  return(publishable_p_pointless == publishable_p)
}

# proportion of congruent conclusions
mean(replicate(10000, simulation()))
```