A method to streamline p-hacking

Ian Hussey

The analytic strategy of p-hacking has rapidly accelerated our advancement of the goals of psychological science (e.g., publications, tenure: Bakker, van Dijk, & Wicherts, 2012), but has suffered a number of setbacks in recent years. In order to remediate this, this article presents a statistical approach that can greatly accelerate and streamline the p-hacking process: generating random numbers that are $< .05$. I refer to this approach as $p_{ointless}$. 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: Munafò et al., 2017), I argue that vast amounts of time and research funding could be saved through the widespread adoption of this innovative statistical approach.

With a few recent and unfortunate exceptions (e.g., Camerer et al., 2018; Klein et al., 2018; Open Science Collaboration, 2015), the discovery that p values can be hacked to support researchers' hypotheses has proven to be of exceptional utility to the enterprise of psychological science (e.g., acquiring publications, tenure, and flair; see Bakker et al., 2012; Simmons, Nelson, & Simonsohn, 2011 for tutorials). However, efforts to further optimize the process of p-hacking have slowed in recent years due to a number of unfortunate setbacks such as wider use of replication and pre-registration (Munafò et al., 2017; Nosek et al., 2015; Nosek, Ebersole, DeHaven, & Mellor, 2018).

In this article, I introduce the $p_{ointless}$ 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 psychological science's primary goals (e.g., high impact publications and tenure). 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). Second, it retains a far more important property of hacked p values: by guaranteeing significant results, it has higher 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 statistics).

Methods and results

The $p_{ointless}$ metric follows the same internal logic as traditional p-hacked analytic strategies (e.g., Bem, 2011). Loosely speaking, this algorithm mimics the outcomes of overt p-hacking behaviour (i.e., exploitation of researcher degrees of freedom until $p < .05$: Simmons et al., 2011). The $p_{ointless}$ metric was inspired by the observation that, regardless of the specifics of any given p-hacking strategy, the product of this process is highlight reliable ($p < .05$). 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 random numbers until one is found that is $< .05$. I refer to this approach as a form of machine learning so as to increase my chance of getting published. R code to calculate $p_{ointless}$ is provided below:

```r
# generate random numbers, stop when < .05
p_ointless <- 1
while (p_ointless >= .05) {
  p_ointless <- runif(n = 1)
}

# print this value
print(paste("p_ointless =", round(p_ointless, 3)))
## [1] "p_ointless = 0.032"
```

Decisions made on the basis of traditional hacked p values and the $p_{ointless}$ metric were then compared in 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 $p_{ointless}$ 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. As such, this minor discrepancy is easily ignored. More importantly, execution time for $p_{ointless}$ is less than one second, whereas traditional p-hacking techniques can take hours or days.

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 $p_{ointless}$ are equivalent, and that the latter requires several orders of magnitude less time and resources to calculate. Academic productivity and more importantly flair can therefore be greatly increased through the widespread adoption of this approach.

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

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-018-0399-z

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... 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

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, Epub ahead of print. https://doi.org/10.1073/pnas.1708274114

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

Appendix: R code for simulation

```
simulation <- function() {

  # simulate publishability of results from econo_p
  # set an inital p value
  p_ointless <- 1

  # generate random values for p, and stop when this value is < .05
  while (p_ointless >= .05) {
    p_ointless <- round(runif(1), 3)
  }

  # decision making
  if(p_ointless < 0.05) {
    publishable_p_ointless = TRUE
  } else {
    publishable_p_ointless = FALSE
  }

  # simulation publishability of results from tradition (hacked) p values
  # p value set to upper bound of observable hacked p values
  p = 0.049

  # decision making
  if(p < 0.05) {
    publishable_p = TRUE
  } else {
    publishable_p = FALSE
  }

  # compare decisions made on the basis of hacked p values vs p_ointless
  return(publishable_p_ointless == publishable_p)

}

# proportion of 10,000 simulated cases where conclusions agree
mean(replicate(10000, simulation()))
## [1] 1
```