

32.4: Questionable Research Practices

A popular book entitled “The Compleat Academic: A Career Guide”, published by the American Psychological Association (Darley, Zanna, and Roediger 2004), aims to provide aspiring researchers with guidance on how to build a career. In a chapter by well-known social psychologist Daryl Bem titled “Writing the Empirical Journal Article”, Bem provides some suggestions about how to write a research paper. Unfortunately, the practices that he suggests are deeply problematic, and have come to be known as *questionable research practices* (QRPs).

Which article should you write? *There are two possible articles you can write: (1) the article you planned to write when you designed your study or (2) the article that makes the most sense now that you have seen the results. They are rarely the same, and the correct answer is (2).*

What Bem suggests here is known as *HARKing* (Hypothesizing After the Results are Known)(Kerr 1998). This might seem innocuous, but is problematic because it allows the research to re-frame a post-hoc conclusion (which we should take with a grain of salt) as an a priori prediction (in which we would have stronger faith). In essence, it allows the researcher to rewrite their theory based on the facts, rather than using the theory to make predictions and then test them – akin to moving the goalpost so that it ends up wherever the ball goes. It thus becomes very difficult to disconfirm incorrect ideas, since the goalpost can always be moved to match the data. Bem continues:

Analyzing data *Examine them from every angle. Analyze the sexes separately. Make up new composite indices. If a datum suggests a new hypothesis, try to find further evidence for it elsewhere in the data. If you see dim traces of interesting patterns, try to reorganize the data to bring them into bolder relief. If there are participants you don't like, or trials, observers, or interviewers who gave you anomalous results, drop them (temporarily). Go on a fishing expedition for something — anything — interesting. No, this is not immoral.*

What Bem suggests here is known as *p-hacking*, which refers to trying many different analyses until one finds a significant result. Bem is correct that if one were to report every analysis done on the data then this approach would not be “immoral”. However, it is rare to see a paper discuss all of the analyses that were performed on a dataset; rather, papers often only present the analyses that *worked* - which usually means that they found a statistically significant result. There are many different ways that one might p-hack:

- Analyze data after every subject, and stop collecting data once $p < .05$
- Analyze many different variables, but only report those with $p < .05$
- Collect many different experimental conditions, but only report those with $p < .05$
- Exclude participants to get $p < .05$
- Transform the data to get $p < .05$

A well-known paper by Simmons, Nelson, and Simonsohn (2011) showed that the use of these kinds of p-hacking strategies could greatly increase the actual false positive rate, resulting in a high number of false positive results.

32.4.1 ESP or QRP?

In 2011, Daryl Bem published an article (Bem 2011) that claimed to have found scientific evidence for extrasensory perception. The article states:

This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by “time-reversing” well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. ...The mean effect size (d) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results.

As researchers began to examine Bem's article, it became clear that he had engaged in all of the QRPs that he had recommended in the chapter discussed above. As Tal Yarkoni pointed out in [a blog post that examined the article](#):

- Sample sizes varied across studies
- Different studies appear to have been lumped together or split apart
- The studies allow many different hypotheses, and it's not clear which were planned in advance
- Bem used one-tailed tests even when it's not clear that there was a directional prediction (so alpha is really 0.1)
- Most of the p-values are very close to 0.05
- It's not clear how many other studies were run but not reported

This page titled [32.4: Questionable Research Practices](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.