

Meredith Turner / Saturday, October 01, 2016 / Categories: 542

I-O Psychology's Lack of Research Integrity

Sheila K. List and Michael A. McDaniel

In recent years, the integrity of our scientific research has been called into question by the popular press who has asked if the scientific method is flawed (Lehrer, 2010). This assertion has been examined by many researchers as well (e.g., Bedeian, Taylor, & Miller, 2010; Kepes & McDaniel, 2013; O'Boyle, Banks, & Gonzalez-Mulé, in press). These authors have argued that the current states of I-O psychology and management are flawed for several reasons. First, the theory fetish (Hambrick, 2007) in our field is making it nearly impossible to publish null results or replications, which has prevented us from developing solid theory (Cucina & McDaniel, 2016). Second, for academics, the necessity to publish for tenure, retention, promotion, raises, and so on encourages researchers to engage in questionable research practices (QRPs) if the obtained results do not align with a priori expectations or do not reach statistical significance (e.g., Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016; O'Boyle et al.).

A recent article in *The Industrial-Organizational Psychologist* by Nicklin, Gibson, and Grand (2016) touched briefly on the prevalence and impact of QRPs on scientific research in its description of two separate panel discussions conducted at the annual conference for the Society for Industrial-Organizational Psychology (SIOP). This article responds to their call to continue the conversation and aims to extend their discussion by reviewing three common QRPs, discussing their prevalence, and outlining the ways in which they undermine our field, and science as a whole. We then discuss several ways which we as a field, and as individual researchers, may discourage QRPs, encourage transparency, and increase the integrity of our results.

What Are QRPs?

QRPs are practices pertaining to analysis of data and reporting of results that may make the results of a study seem more favorable and have the potential to mislead consumers of the research (Banks, Rogelberg et al., 2016). QRPs include such practices as removing or adding data after testing hypotheses, adding or removing control variables after hypothesis testing, and altering, adding, or removing hypotheses after testing. Although each of these practices may be used to intentionally provide misleading results, it is also possible for authors to engage in these practices for less insidious reasons. Consider this situation. A researcher drops data from two participants who were identified as outliers, leading to a statistically significant result not previously observed, which could be viewed as inappropriate. However, if the researcher describes the process by which the data were identified as outliers and states that the two cases were dropped, most researchers would not consider the practice unethical. Thus, determining what is a QRP and what is not is, at least partially, a question of transparency (Fanelli, 2013).

Several authors have proposed taxonomies of QRPs including Banks, Rogelberg et al. (2016) and O'Boyle et al. (in press). Ultimately, however, most QRPs can be considered as some form of HARKing (hypothesizing after results are known; Kerr, 1998), selective reporting, or p-hacking. Kerr defined HARKing as presenting post-hoc hypotheses as a priori hypotheses. Kerr outlined some initial evidence indicating that HARKing may be widespread in the field of psychology. More recent evidence has also demonstrated the very common prevalence of HARKing. For instance, Bedeian et al. (2010) found that 91.9% of the 384 faculty surveyed stated that they had knowledge of faculty engaging in HARKing within the past year. Other estimates have suggested that between approximately 27% and 90% of researchers have engaged in HARKing, or know of another faculty member who has (John, Loewenstein, & Prelec, 2012).

Selective reporting occurs at both the hypothesis level and the study level. For instance, if a researcher conducted a study in which half of the hypotheses were not supported by the data, the researcher may choose to report only the hypotheses that were supported. Alternatively, if a researcher conducts a study where none of the central hypotheses are reported, that researcher may abandon the entire study and not submit it to a journal. Approximately 46% to 50% of researchers surveyed admitted to selectively reporting hypotheses (John et al., 2012). Furthermore, Bedeian et al. (2010) found that approximately 50% of faculty members surveyed stated that they knew someone who had withheld data that contradicted a previous finding (another motivation for selective reporting). Many would find it a reasonable inference that self-reports of such practices substantially underestimate the frequency of occurrence.

In addition to self-report survey evidence regarding the prevalence of HARKing and selective reporting, researchers have also examined the way in which hypotheses change from dissertations and conference papers to published articles (e.g., Banks, O'Boyle, White, & Batchelor, 2013; Mazzola & Deuling, 2013; O'Boyle et al., in press). These researchers have found that the number of hypotheses supported in journal articles is significantly greater than those supported in dissertations and conference papers. This suggests that the authors of these papers are engaging in selective reporting and/or HARKing and that these practices are common.

In addition to HARKing and selective reporting, the examination of the instances and results of p-hacking has also increased (e.g., Field, Baker, Bosco, McDaniel, & Kepes, 2016). P-hacking is a broader category of QRPs that includes any practice that a researcher uses to turn a nonstatistically result into a statistically significant one. This can include collecting data until significance is achieved, ceasing the collection of data once significance is achieved, and adding or deleting control variables based on which variables result in a significant relation (Head, Holman, Lanfear, Kahn, & Jennions, 2015). Evidence based on the expected probability of p values of less than .05 has shown that p-hacking is common across many disciplines, including psychology (e.g., Head et al.; Masicampo & Lalande, 2012). Furthermore, survey evidence has demonstrated that approximately 56% or more of researchers have either admitted continuing to collect data after testing their hypotheses or know of a researcher who has done so (John et al., 2012). Furthermore, 22-23% of researchers admitted to a different form of p-hacking, rounding down p -values to .05 when they are, in fact, greater than .05 (e.g., .054; John et al. 2012).

Do QRPs Matter?

As outlined above, research has shown that many QRPs occur frequently. However, is it possible that they do not hurt the scientific process? The mission of many research universities involves the creation and dissemination of knowledge; however, if the research in our field is judged to be untrustworthy, we may ultimately be failing at these endeavors. In regard to the creation of knowledge, the prevalence of QRPs makes it difficult to sort out the true relations between variables. Kepes and McDaniel (2013) discussed how problems in the publication and reporting process, including QRPs, distort scientific knowledge through their impact on meta-analyses. For example, the selective reporting of hypotheses and lack of outlets for null results may lead to over-inflated meta-analytic correlations. Given that meta-analyses synthesize the results of several primary studies to examine the extent to which hypotheses are supported, they are generally more accurate than any one primary study. However, their accuracy is damaged by publication bias, which is the extent that the studies included in a meta-analysis are not representative of all studies conducted that evaluate the relation of interest (Banks, Kepes, & McDaniel, 2012; Kepes & McDaniel, 2015; Rothstein, Sutton, & Borenstein, 2005). Ferguson and Brannick (2012) found that approximately 40% of meta-analyses in psychology are affected by publication bias, and approximately 25% of meta-analyses have a worrisome degree of publication bias. Evidence suggests that in the social sciences, the main contributor to publication bias is the selective reporting of hypotheses and the suppression of null findings (Franco, Malhotra, & Simonovits, 2014). Recent reexaminations of well-established meta-analytic correlations have found that these relations may be overestimated (e.g., Kepes & McDaniel, 2015; Renkewitz, Fuchs, & Fiedler, 2011).

In addition to the effect of publication bias, HARKing has also been shown to impact the effect sizes through inflation. Bosco, Aguinis, Field, Pierce, and Dalton (2015) compared cumulative effect sizes from articles that explicitly hypothesized relations to cumulative effect sizes from articles whose relations were included but not explicitly hypothesized (i.e., the relation could be determined through examining the correlation table, but there was not a specific hypothesis about the relation between the variables in the article). They found that the effect sizes from studies with hypothesized relations were significantly larger than those where there was not a hypothesis. One could infer that when relations have smaller effect sizes, or are not statistically significant, the authors engage in HARKing so as to have fewer unsupported hypotheses (Bosco et al.). The influence of p-hacking has also been examined. However, the evidence regarding whether or not it affects the quality of scientific findings is mixed. Head et al. (2015) found that p-hacking did not significantly impact the quality of meta-analytic evidence, whereas other researchers purport that the prevalence of p-hacking suggests that most of the published research findings are inaccurate (Ioannidis, 2005).

In regards to the dissemination of knowledge, QRPs harm the reputation of science for the lay population. As Bedeian et al. (2010) discussed, researchers who engage in, or even appear to engage in, QRPs that flout the basic tenets of science make it harder for lay individuals to take scientific findings seriously. If scientific research is untrustworthy, then the ability to disseminate useful knowledge is diluted. Furthermore, the untrustworthiness of research makes it difficult to practice evidence-based management (Kepes, Bennett, & McDaniel, 2014), as we cannot be confident in the accuracy of the evidence. Last, for those psychologists who train graduate students, it is easy to forget that graduate students not only learn through coursework and conducting their own research, but they also learn by observing how their mentors conduct research. These lessons can be conflicting. Banks, Rogelberg et al. (2016) found that graduate students reported receiving instructions to avoid engaging in QRPs but also frequently observed them. These behaviors become habit. Graduate students learn that although in a perfect world, QRPs should be avoided; in practice, most everyone engages in them. In fact, the "publish or perish" mantra, coupled with the research findings that number of top-tier publications is the main predictor of salary (Gomez-Mejia & Balkin, 1992), socializes graduate students that it is acceptable to engage in QRPs because that is how they were trained and how they will be incentivized.

Thus, ultimately QRPs matter for three main reasons: QRPs undermine our cumulative knowledge, damage our integrity, and provide a poor example for future researchers. Therefore, QRPs are a stain on our field, regardless of whether the individuals engaging in them have mal-intent or not.

How Can We Promote Research Integrity?

Several different researchers have offered recommendations aimed at reducing the frequency of QRPs and increasing the integrity of our research (e.g., Banks, O'Boyle et al., 2016; Banks, Rogelberg et al., 2016; Kepes et al., 2014; Kepes et al., 2013; O'Boyle et al., in press). These recommendations include clarifying what is considered acceptable and what is considered unacceptable, encouraging the publication of exploratory studies, replications, and null findings, and changing the review process. Although each of these are useful suggestions that we feel should be implemented, we focus here specifically on what individual researchers can do to improve the integrity of their studies.

First, we concur with Nicklin et al. (2016) and many other authors that researchers should emphasize transparency. This means not only conforming to the Journal Articles Reporting Standards (JARS) of the American Psychological Association (APA, 2008) and the Meta-Analytic Reporting Standards (MARS; APA, 2008) by making a concerted effort to write clearly and include all relevant information in journal articles but also sharing data and any code or syntax used to analyze data. Data and syntax can be made available through the Open Science Framework (<http://osf.io>) and included in supplemental article materials maintained by most journals. Ultimately this critical appraisal of research improves the confidence that we can have in the results by allowing other researchers to reanalyze data using different techniques and confirm findings, making the findings more impactful (Baker, Bosco, Uggerslev, & Steel, 2016). Furthermore, this critical appraisal process encourages continued professional development among researchers.

Second, researchers should perform and report sensitivity analyses on their results to determine the robustness of their conclusions. Sensitivity analyses provide supplementary evidence that allow more confidence to be placed in the results of a study. For example, authors can analyze data using different controls, different measures or operationalizations of the variables of interest, different analytical methods, and so forth. Including these checks ensures that the effects observed in the study is not just a function of the particular combination of variables and measures, or the data analysis technique employed, but rather is a reflection of the true underlying effect. Reporting the sensitivity analyses enhances the transparency of the research and credibility of the conclusions. This is hardly ever done in I-O psychology.

Researchers conducting meta-analyses should also be sure to conduct sensitivity analyses. In the meta-analytic context, sensitivity analyses include the assessment of publication bias and outlier analysis. Examining the extent to which publication bias impacts meta-analytic correlations is extremely important given the prevalence assessment of publication bias completed by Ferguson and Brannick (2012). However, Banks, Kepes, and McDaniel (2012) found that only 31% of meta-analyses published in top management and I-O psychology journals empirically assessed the presence of publication bias. Furthermore, many of the articles that did include an assessment of publication bias used inappropriate methods (Banks et al., 2012). In regards to outliers, despite the knowledge that outliers can impact meta-analytic conclusions, only about 3% of meta-analytic studies include an assessment of outliers (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011).

There are more than 10 different methods to assess publication bias, which are each affected differently by different factors (e.g., number of studies included, degree of heterogeneity present, assumptions of symmetry). There are also multiple methods for outlier detection (Viechtbauer & Cheung, 2010). Currently, there is no accepted standard regarding which method is the best to use in every scenario; thus, it is important to employ multiple methods and take a triangulation approach. Using multiple methods will allow meta-analysts to estimate the mean correlation more accurately (Kepes, Banks, McDaniel, & Whetzel, 2012; Kepes & McDaniel, 2015).

Nicklin et al.'s (2016) review of topics discussed at the annual SIOP conference provided a jumping off point for our review. Clearly, these issues present challenges that we must face and tackle in order to continually improve as a field. It is encouraging to know that we as a field are engaging in meaningful conversations aimed at arising to these challenges. Yet, although it is important to discuss the importance of research integrity and how it can be improved upon, it is essential that we do more than talk. It is time to put recommendations mentioned Nicklin et al. (2016), Banks, Rogelberg et al. (2016), Kepes et al. (2013), O'Boyle et al. (in press), and by many other authors, into practice. The QRP research in I-O psychology is at the crossroads (Johnson, 1936). Research and the popular press has addressed the prevalence of QRPs which undermine the trustworthiness of our findings and our integrity as researchers. Now, we as a field must determine what to do with this knowledge. Do we wish to be known as untrustworthy researchers, or as scholars who promote the advancement of science through adherence to practices consistent with research integrity? Time will tell.

References

- Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, 37, 5-38. doi: 10.1177/0149206310377113
- American Psychological Association (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, 839-851. doi: 10.1037/0003-066X.63.9.839

- Baker, C. A., Bosco, F. A., Uggerslev, K. L., & Steel, P. G. (2016). metaBUS: An open search engine of I-O research findings. *The Industrial-Organizational Psychologist*
- Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*, 20, 182-196. doi:10.1111/j.1468-2389.2012.00591.x
- Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., . . . Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, 42, 5-20. doi: 10.1177/0149206315619011
- Banks, G. C., O'Boyle, E. H., White, C. D., & Batchelor, J. H. (2013). *Tracking SMA papers to journal publication: An investigation into the phases of dissemination bias*. Paper presented at the 2013 annual meeting of the Southern Management Association, New Orleans, LA.
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, 1-16. doi: 10.1007/s10869-016-9456-7
- Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education*, 9, 715-725. doi: 10.5465/amle.2010.56659889
- Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*, 69, 709-750. doi:10.1111/peps.12111
- Cucina, J.M., & McDaniel, M. A. (2016). Pseudosocial theory proliferation is damaging the organizational sciences. *Journal of Organizational Behavior*. doi: 10.1002/job.2117
- Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature*, 494, 149. doi: 10.1038/494149a
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17, 120-128. doi: 10.1037/a0024445
- Field, J. G., Baker, C. A., Bosco, F. A., McDaniel, M. A., & Kepes, S. (2016, April). *The extent of p-hacking in I-O psychology*. Paper presented at the 31st Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505. doi: 10.1126/science.1255484
- Gomez-Mejia, L. R., & Balkin, D. B. (1992) Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, 35, 921-955. doi: 10.2307/256535
- Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, 50, 1348 -1352
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, 13, doi: 10.1371/journal.pbio.1002106
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *CHANCE*, 18(4), 40-47. doi:10.1080/09332480.2005.10722754
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524-532. doi: 10.1177/0956797611430953
- Johnson, R. (1936). *Crossroad Blues*. Brunswick Records: San Antonio. Available at <https://www.youtube.com/watch?v=Yd60nl4sa9A>
- Kepes, S., Banks, G.C., McDaniel, M.A., & Whetzel, D.L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15, 624-662. doi: 10.1177/1094428112452760
- Kepes, S., Bennett, A. A., & McDaniel, M. A. (2014). Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education*, 13,

446-466. doi: 10.5465/amle.2013.0193

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

Kepes, S. & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS ONE*, 10(10):e0141468. doi: 10.1371/journal.pone.0141468.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196-217. doi: 10.1207/s15327957pspr0203_4

Lehrer, J. (2010). The truth wears off. *The New Yorker*, 86, 53-57.

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of *p* values just below .05. *Quarterly Journal of Experimental Psychology*, 65, 2271-2279. doi: 10.1080/17470218.2012.711335

Mazzola, J. J., & Deuling, J. K. (2013). Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I-O journal articles. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 6, 279-284. doi: 10.1111/iops.12049

Nicklin, J. M., Gibson, J. L., & Grand, J. (2016). Cultivating a future of meaningful, impactful, and transparent research. *The Industrial-Organizational Psychologist*

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mule, E. (in press). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*. doi: 10.1177/0149206314527133

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication bias in JDM research? *Judgment and Decision Making*, 6, 870-881.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. West Sussex, UK: Wiley.

Viechtbauer, W. and Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 112-125.

320 Rate this article: No rating