Industrial and Organizational Psychology (2022), **15**, 533–536
doi:10.1017/iop.2022.68

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

Revisiting the paradox of replication: Is the solution to the paradox big data style research or something else?

In-Sue Oh

Department of Management, Fox School of Business, Temple University, Philadelphia, PA 19122, USA
Email: insue.oh@temple.edu

In the focal article, Guzzo et al. (2022) identify and discuss several incompatibilities between the promises and principles of the recent open science movement and the prospects of increased research credibility and scientific progress via the movement. One of them is what they term the paradox of replication—the pursuit of replication in the open science movement will result in less robust and credible research findings by unintendedly allowing and encouraging researchers to conduct and publish more small-sample exact replications with more risk of inflated effect sizes. They also argue that big data style research, as a solution to the paradox, can yield more robust and credible research findings, thus contributing to increased research credibility and scientific progress, the ultimate mission of open science. Although I see some merits in their arguments, I also see some limitations and needs for further elaboration.

First, Guzzo et al. (2022) argue that "open science's emphasis on replication to attempt to assure the certainty of published findings will encourage projects with small sample sizes that should necessarily raise concerns of adequate statistical power" (p. 23). It is widely known that the typical sample size in published articles has not sufficiently increased over the years, despite repeated warnings about the law of small numbers (Tversky & Kahneman, 1971), the erroneous belief that "small random samples can be considered to be about as representative of their corresponding populations as large random samples are" (Schmidt & Hunter, 1978, p. 216). In fact, median sample sizes in the articles published in the *Journal of Applied Psychology* during the 1995–2008 period have been consistent at around 173 with a slight increase over the 13-year period examined (Shen et al., 2013). Moreover, journals have been supportive of publishing small-sample (replication) studies at least implicitly for decades, given a large number of meta-analyses currently available that are based on such replication studies. For example, Judge et al. (2001) meta-analysis on the relationships between job satisfaction and job performance is based on 312 independent samples (replications) with an average sample size of 175, a relatively small size in Guzzo et al.'s view. Thus, although it remains to be seen whether Guzzo et al.'s prediction above is accurate, it is difficult to see how the open science movement will lead to an even larger number of published small-sample (replication) studies and ultimately to a decrease in sample size in journal articles, given the historical trend that does not support their prediction.

Second, Guzzo et al. (2022) argue that the recent open science movement will result in "an overvaluing of exact replication of specific findings" (p. 23). To better understand this argument, one needs to know that there are different forms of replications. According to a typology by Kelly et al. (1979), there are four different types of replications: (a) a literal (as known as exact, strict) replication is a new study conducted using the same independent and dependent variables in the

I would like to thank the late Frank Schmidt for many ideas in this article.

original study; some even defined it more narrowly as a new study conducted by the same research team in exactly the same way as in the original study (Aronson et al., 1990); (b) an operational replication is a new study conducted using the same independent variable in the original study, but with a different operationalization of the dependent variable; (c) an instrumental replication is a new study using the same dependent variable in the original study, but with a different operationalization of the independent variable, and (d) a constructive (also known as systematic) replication is a new study with different operationalizations of both independent and dependent variables. After surveying three issues of the *Journal of Personality and Social Psychology* published in 1993, Neuliep & Crandall (1993) found that 78% of studies examined were replications in one form or another, although never literal. If what Guzzo et al. mean by "exact" replication is the literal replication mentioned above, it is difficult to see why and how journals suddenly publish more literal replication studies than before due to the recent open science movement, given Neuliep and Crandall's discovery. If what Guzzo et al. mean by "exact" replication is either the operational or instrumental replication mentioned above, the truth is that journals have already valued such studies. For example, most if not all studies included in Judge et al. (2001) meta-analysis mentioned above are either instrumental or operational replications.

Third, Guzzo et al. (2022) argue that "a literature of replicated small-sample studies runs the risk of overestimating the magnitude of 'true' relationships" (pp. 23–24). I see merit in their argument given their data-based rationale: Kűhberger et al. (2014) reported a meta-analytic correlation of -.54 between sample size and effect size based on about 400 studies published in various psychology journals. In my view, equally, if not more, important findings in Kűhberger et al.'s article (not mentioned by Guzzo et al.) are that the ratio of published studies with *p*-values just below .05 (just significant) to those with *p*-values just above .05 (just non-significant) is about 3:1. As discussed in Schmidt & Oh (2016), "given typical effects sizes and typical *N*s, and the resulting typical levels of statistical power, then in the absence of publication bias and questionable research practices, it is clear that about half of all studies will report significant findings and half will report nonsignificant findings" (p. 34). Then, the natural follow-up question is: what causes the unlikely ratio of 3:1 instead of a more likely ratio of 1:1? Many scholars have reasoned that this tends to occur when researchers use questionable research practices, such as *p*-hacking (see Schmidt & Hunter, 2015, pp. 513–553). As such, the real problem underlying the higher (inflated) effect sizes in smaller-sample studies is not necessarily the "small" sample size itself but more likely questionable research practices.

Fourth, Guzzo et al. (2022) argue that big data style research could address problems associated with the paradox of replication. For example, they state that "compared to its small-sample counterparts big data style research can yield more robust findings along with less risk of effect size overestimation" (p. 24). It makes sense that big data and, by extension, larger-sample studies are more likely to (or be condoned to) report smaller effect sizes than smaller-sample studies because nearly all relationships will be statistically significant, and significant (vs. non-significant) findings are more likely to be published. However, there are several noteworthy issues Guzzo et al. fail to discuss in the focal article that could seriously limit the superiority of big data style research over meta-analysis as a solution to the paradox of replications (see Schmidt & Hunter, 2015, pp. 372–374). To begin with, most researchers do not have access to resources (e.g., money and time) necessary to conduct big data or large-sample studies. That is, a more feasible solution to achieving sufficient statistical power is to conduct multiple small-sample studies (preferably in the form of constructive replication) and later synthesize them via meta-analysis. Moreover, big data style research often prioritizes sample size at the expense of other important aspects of research. For example, in survey research, many large-sample studies currently available often use single-item or brief measures that are less psychometrically sound (less reliable and construct-valid) than desirable. In experimental research, quasi-experiments (in which random assignments of subjects are not possible, thus leading to low internal validity) tend to be based on larger samples than well-controlled lab experiments. In addition, participants in many large-sample

https://doi.org/10.1017/iop.2022.68 Published online by Cambridge University Press

survey-based studies currently available are often based on a specific segment of the population. These studies themselves cannot guarantee whether their findings are generalizable to the general workforce population, the target population in most applied psychological research in work settings. Taken together, we cannot completely rule out the possibility that the lower effect sizes associated with larger-sample (big data) studies are partly attributable to their suboptimal study characteristics mentioned above, such as lower scale reliability.

Fifth, and related to the point above, I agree with Guzzo et al. (pp. 31–32) that a more practical solution to and achieving more robust and credible research findings is to encourage "constructive" replications regardless of whether the findings are statistically significant or not. This is consistent with Eden (2002) statement that "the *less* similar a replication is to an original study, the greater its potential contribution" (p. 842). From a triangulation perspective, if constructive replications lead to the same conclusion, it increases external validity and research credibility. This naturally begs the question: how can one determine whether constructive replications lead to the same conclusion? Although Guzzo et al. assert that big data style research plays a significant role here by arguing that "conceptual (constructive) replication opportunities abound in big-data research" (p. 33, parenthesis added for clarity), it is rather difficult to see how big data style research enables ample constructive replications. I believe this is where meta-analysis comes into play instead. Specifically, meta-analysis allows us to gauge how much of the variation in effect size across input (replication) studies is real or artifactual due to sampling error and other artifacts. Suppose all or most of the variation across the studies is artifactual. In that case, one can conclude that the relationship in question is practically the same across varied replications (e.g., differences in operationalization of study variables). If little of the variation is artifactual, the conclusion is that there exist some moderators, and one should search for them (Oh & Roth, 2017). This type of information is not available in big data style research. That is, "A single large study with sample size equal to the sum of the study sample sizes in such a meta-analysis is not capable of revealing these facts" (Schmidt & Oh, 2016, p. 33). Some argued that such information (true between-studies heterogeneity) could be obtained by dividing a large sample study into many smaller samples. Guzzo et al. (2022) also discuss a similar idea: "The presence of big data … enable[s] multiple replications and/or meta-analysis processes within a single study's large data set" (p. 16). However, the heterogeneity estimate obtained via meta-analysis in this way is not informative because such small studies are not only literal replications but also not statistically independent of each other. Besides, if multiple big data or large-sample studies are available on the same relationship, we still need meta-analysis to synthesize these studies to draw more accurate and informative conclusions (Schmidt & Hunter, 2015).

In conclusion, Guzzo et al. (2022), in the focal article, argue that the recent open science movement may unintendedly result in more (a) "small-sample," (b) "exact replication," and (c) "questionable" (with inflated effect sizes) studies, thus leading to the paradox of replication—the emphasis on replication, paradoxically, results in a decrease, rather than an increase, in more robust and credible research findings. They further argue that big data style research is a solution to the paradox. However, as discussed above, I believe, like the power of *small wins*, a better solution to the paradox is a large number of small-sample (constructive replication) studies which combined via meta-analysis will contribute more to research credibility and scientific progress, the ultimate mission of open science.

References

Aronson, E., Ellsworth, P., & Gonzales, M. (1990). *Methods of research in social psychology* (2nd Ed.). New York: McGraw-Hill.

Eden, D. (2002). Replication, meta-analysis, scientific progress, and *AMJ*'s publication policy. *Academy of Management Journal*, **45**(5), 841–846. https://www.jstor.org/stable/3069317

https://doi.org/10.1017/iop.2022.68 Published online by Cambridge University Press

Guzzo, R. A., Schneider, B., & Nalbantian, H. R. (2022). Open science, closed doors: The perils and potential of open science for research in practice. *Industrial and Organizational Psychology*, **15**(4), 495–515.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, **127**(3), 376–407. https://doi.org/10.1037/0033-2909.127.3.376

Kelly, C. W., Chase, L. J., & Tucker, R. K. (1979). Replication in experimental communication research: An analysis. *Human Communication Research*, **5**, 338–342. https://doi.org/10.1111/j.1468-2958.1979.tb00646.x

Kűhberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE*, **9**(9): e105825. doi: 10.1371/journal.pone.0105825

Neuliep, J. W., & Crandall, R. (1993). Everyone was wrong: There are lots of replications out there. *Journal of Social Behavior and Personality*, **8**, 1–8.

Oh, I.-S., & Roth, P. L. (2017). On the mystery (or myth) of challenging principles and methods of validity generalization (VG) based on fragmentary knowledge and improper or outdated practices of VG. *Industrial and Organizational Psychology*, **10**(3), 479–485. https://doi.org/10.1017/iop.2017.45

Schmidt, F. L., & Hunter, J. E. (1978). Moderator research and the law of small numbers. *Personnel Psychology*, **31**(2), 215–232. https://doi.org/10.1111/j.1744-6570.1978.tb00441.x

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, **4**(1), 32–37. https://doi.org/10.1037/arc0000029

Shen, W., Kiger, W., Davies, T. B., Rasch, S. E., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology*, **96**(5), 1055–1064. https://doi.org/10.1037/a0023322

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, **76**(2), 105–110. https://doi.org/10.1037/h0031322

Cite this article: Oh, I-S. (2022). Revisiting the paradox of replication: Is the solution to the paradox big data style research or something else? *Industrial and Organizational Psychology* **15**, 533–536. https://doi.org/10.1017/iop.2022.68