Published in final edited form as:

Int J Psychophysiol. 2020 August ; 154: 3–5. doi:10.1016/j.ijpsycho.2019.05.006.

What the replication crisis means for intervention science

Frank G. Hillary[a,b,*], **John D. Medaglia**[c,d,e]

[a]Penn State Department of Psychology, Penn State University, United States of America

[b]Social, Life, and Engineering Imaging Center (SLEIC), United States of America

[c]Department of Psychology, Drexel University, United States of America

[d]Department of Neurology, Drexel University, United States of America

[e]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, United States of America

Abstract

The provocative paper by Ioannidis (2005) claiming that "most research findings are false" re-ignited longstanding concerns (see Meehl, 1967) that findings in the behavioral sciences are unlikely to be replicated. Then, a landmark paper by Nosek et al. (2015a) substantiated this conjecture, showing that, study reproducibility in psychology hovers at 40%. With the unfortunate failure of clinical trials in brain injury and other neurological disorders, it may be time to reconsider approaches not only in clinical interventions, but also how we establish their efficacy. A scientific community galvanized by a history of failed clinical trials and motivated by this "crisis" may be at critical cross-roads for change engendering a culture of transparent, open science where the primary goal is to test and not support hypotheses about specific interventions. The outcome of this scientific introspection could be a paradigm shift that accelerates our science bringing investigators closer to important advancements in rehabilitation medicine. In this commentary we offer a brief summary of how open science, study pre-registration and reorganization of scientific incentive structure could advance the clinical sciences.

Can the "Crisis" evoke a Paradigm Shift in Intervention Science?

The provocative paper by Ioannidis (2005) claiming that "most research findings are false" re-ignited longstanding concerns (see Meehl, 1967) that findings in the behavioral sciences are unlikely to be replicated. Then, a landmark paper by Nosek et al. (2015a) substantiated this conjecture, showing that reproducibility hovers at 40%. While one might argue about the terminology and causes used to describe the situation facing behavioral scientists ("replication crisis"; cf. Maxwell et al., 2015), there is growing consensus that there is room for improvement in approach and methods used and these concerns have spared few areas of research in the health sciences (Benjamin et al., 2018; Friesike et al., 2015; Goodman et al., 2016; Munafò et al., 2017; Nosek et al., 2015b).

*Corresponding author at: 313 Bruce V. Moore Bldg., University Park, PA 16801, United States of America. fhillary@psu.edu (F.G. Hillary).

Central to concerns in the replication crisis is a demand for larger data sets for maximizing statistical power while simultaneously questioning the incentive structure for publishing only statistically significant findings (Cohen, 2016) and adhering to the philosophically flawed null-hypothesis significance testing (NHST) (see Henkel, 2017; Schneider, 2015). One might argue that the goals of rehabilitation medicine and, cognitive remediation specifically, sit at an unenviable intersection occupied by studies with small(ish) sample sizes aiming to detect often subtle effects buried in the noise of inter and intra subject variability (Park and Ingles, 2001; Rohling et al., 2009; Sitzer et al., 2006; Wykes et al., 2011). To make matters worse, the goal to demonstrate statistically significant effects in interventions must be achieved while surpassing increasingly stringent statistical thresholds that have been proposed to handle replication problems based in NHST (Benjamin et al., 2018).

With the unfortunate failure of clinical trials in brain injury and other neurological disorders, it may be time to reconsider approaches not only in our interventions, but also how we establish their efficacy. In the case of pharmacologic interventions in acute traumatic brain injury (TBI) in particular, the staggering 100% rate of failure (see Stein, 2015) has left the community to ponder how so many promising interventions could survive early studies, only to falter so impressively during phase III clinical trials (see Menon & Maas, 2014). To understand where things have gone wrong with both behavioral and pharmacologic interventions, one place to look would be the structures in place guiding how we set-out to study the phenomenon in the first place.

The “replication crisis” not only highlights the limitations of traditional statistical approaches and the circumscribed requirements for scientific publication, but it leads to questions about the culture of science. The culture of medical science includes an incentive structure that requires innovative approaches, novel findings, and validation through statistical significance via NHST. If NHST fails, researchers commonly test many post-hoc hypotheses in order to fit the data (i.e., p-hacking; Head et al., 2015). Unfortunately, this culture does not promote efficient science or the open study of clinical research because researchers are not incentivized to publish or share results with the scientific community when interventions fail. It is essential for the scientific community to be aware of both successes and failures of well-designed clinical interventions, making null findings a vital part of the scientific landscape and ultimately expediting research.

There is also an important need to understand how group data from an intervention study can inform us about the efficacy of any intervention in the individual. Drawing from ergodic theory, Molenaar (2004) predicted that cases where statistical estimates based on group data would rarely reflect processes within individuals. In fact, empirical studies show that individual and group estimates do diverge considerably (Fisher et al., 2018; Seghier and Price, 2018). In the worst case, we would need a different mechanistic model for each person to treat that person’s cognitive deficit or disease process. In other words, it is unclear how to understand and treat cognitive dysfunction with group data without knowing how group-level inferences map into individual processes over time. Thus, directly measuring within-subject variability is a central feature to precision medicine to determine which failures to replicate are driven by a lack of person-level analysis. In rehabilitation medicine,

reproducibility is at least partially linked to how well group-level data represent individual responses to treatment. Therefore, the person-level is the appropriate level at which interventions should be conceptualized and studied and our group-level claims should be rooted in models of processes that are validated within persons. These processes could be tested for validity with convergent behavioral, neuroimaging, and other biological measures to identify mechanisms of action that could target precision interventions. In principle, starting with the individual as the basic unit of study could reduce the time needed to discover mechanisms of disorders and change in clinical samples.

What to do?

So where do these challenges leave the intervention scientist? While the problems with NHST have been described for decades (see Cohen, 1994; Rozeboom, 1960), the advent of the "crisis" may bring new perspective and a collective imperative to change how we collect, analyze, and disseminate clinical research findings. A general solution to many of the concerns expressed here is to advocate for transparent and open science. One vital step in an open science landscape is study pre-registration, where interventionists using established methods for even small sample or single-subject designs (e.g., multiple baseline, reversal designs) can register the study goals and the results can be accessed by the scientific community regardless of the study outcome. Because all findings are available to the scientific community, open science allows us to eliminate ineffective interventions and aids in identifying interventions that work through replication. Moreover, providing access to the outcomes of all studies can allow investigators to view treatment effects and their variances as a continuum. This could be one way to mitigate the community's reliance on arbitrary NHST criteria and the tendency to artificially declare some findings significant and others not (Wasserstein et al., 2019). As others have begun to note more urgently, scientists should accept a statistical and epistemic worldview that embraces uncertainty at its foundation (McShane et al., 2019).

Perhaps unsurprisingly – and encouragingly, recent analysis of pre-registered studies revealed a sharp rise in the publication of null findings and replication studies (Warren, 2018). Because of this, pre-registration will also reduce the tendency for interventions to appear successful largely because they have been propped-up by well-meaning, but naturally biased, researchers who have been incentivized to defend interventions as opposed to critically testing their efficacy. Open science and study preregistration may also help to standardize methods, which are currently lacking in some areas of the clinical neurosciences including functional brain imaging (see Esteban et al., 2019; Hallquist and Hillary, 2018). Moreover, data sharing fostered in open science holds additional opportunities to test the reliability of interventions and their generalizability between research labs. Open science initiatives can ultimately lead to data repositories that permit estimates for population, sample level, and person level effect sizes. By extension, data sharing provides estimates of patient response distributions that can be used as "priors" for testing a range of hypotheses (including the null) within a Bayesian framework which is becoming increasingly accessible for statistical analyses (e.g., Bayes factor estimates) (Hoijtink et al., 2019). This effort should consider the extent to which aggregate data represent any process within the individuals within groups to quantify potential process variability, clarify mechanisms, and

tailor treatments. Finally, for the goals presented here to be realized it requires a change in scientific culture so that researchers are awarded and promoted based upon their dedication to support open science, data sharing, and study replication, (Gernsbacher, 2018).

One goal of this *Special Issue* to address current challenges in rehabilitation medicine. It is an important time for clinical interventionists to have this conversation. The concerns outlined above with regard to NHST and scientific incentive structure are certainly not new and are not the sole reason for difficulties advancing rehabilitation medicine. However, a scientific community galvanized by a history of failed clinical trials and motivated by this "crisis" may be at critical cross-roads for change engendering a culture of transparent, open science where the primary goal is to test and not support hypotheses about specific interventions. The outcome of this scientific introspection might be a paradigm shift that accelerates our science bringing investigators closer to important advances in rehabilitation medicine.

References

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, ... Cesarini D, 2018 Redefine statistical significance. Nat. Hum. Behav 2 (1), 6. [PubMed: 30980045]

Cohen J, 1994 The earth is round (p < .05). Am. Psychol 49, 997–1003.

Cohen J (2016). The earth is round (p < .05). In What if there were no significance tests? (pp. 69–82). Routledge.

Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ, 2019 fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16 (1), 111–116. 10.1038/s41592-018-0235-4. [PubMed: 30532080]

Fisher AJ, Medaglia JD, Jeronimus BF, 2018 Lack of group-to-individual generalizability is a threat to human subjects research. Proc. Natl. Acad. Sci. U. S. A 115 (27), E6106–E6115. 10.1073/pnas.1711978115. Epub 2018 Jun 18. [PubMed: 29915059]

Friesike S, Widenmayer B, Gassmann O, Schildhauer T, 2015 Opening science: towards an agenda of open science in academia and industry. J. Technol. Transf 40 (4), 581–601.

Gernsbacher MA, 2018 Writing empirical articles: transparency, reproducibility, clarity, and memorability. Advanced Methods and Practice in Psychological Science 1 (3), 403–414.

Goodman SN, Fanelli D, & Ioannidis JP (2016). What does research reproducibility mean?. Science translational medicine, 8(341), 341ps12–341ps12.

Hallquist MN, Hillary FG, 2018 Graph theory approaches to functional network organization in brain disorders: a critique for a brave new small-world. Network Neuroscience 3 (1), 1–26. [PubMed: 30793071]

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD, 2015 The extent and consequences of p-hacking in science. PLoS Biol 13 (3), e1002106. [PubMed: 25768323]

Henkel RE, 2017 The Significance Test Controversy: A Reader Routledge.

Hoijtink H, Mulder J, van Lissa C, Gu X, 2019 A tutorial on testing hypotheses using the Bayes factor. Psychol. Methods 10.1037/met0000201.

Ioannidis JP, 2005 Why most published research findings are false. PLoS Med 2 (8), e124 Epub 2005 Aug 30. [PubMed: 16060722]

Maxwell SE, Lau MY, Howard GS, 2015 Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? Am. Psychol 70 (6), 487. [PubMed: 26348332]

McShane BB, Gal D, Gelman A, Robert C, & Tackett JL (2019). Abandon statistical significance. The American Statistician, 73(sup1), 235–245.

Meehl PE, 1967 Theory-Testing in Psychology and Physics: A Methodological Paradox. Philos. Sci 34 (2), 103–115.

Molenaar PC, 2004 A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. Measurement 2 (4), 201–218.

Munafò MR, Nosek BA, Bishop DV, Button KS, Chambers CD, du Sert NP, ... Ioannidis JP, 2017 A manifesto for reproducible science. Nat. Hum. Behav 1 (1), 0021.

Nosek BA et al. (2015a). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. [PubMed: 26315443]

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, ... Contestabile M, 2015b Promoting an open research culture. Science 348 (6242), 1422–1425. [PubMed: 26113702]

Park NW, Ingles JL, 2001 Effectiveness of attention rehabilitation after an acquired brain injury: a meta-analysis. Neuropsychology 15 (2), 199. [PubMed: 11324863]

Rohling ML, Faust ME, Beverly B, & Demakis G (2009). Effectiveness of cognitive rehabilitation following acquired brain injury: a meta-analytic re-examination of Cicerone et al.'s (2000, 2005) systematic reviews. Neuropsychology, 23(1), 20. [PubMed: 19210030]

Rozeboom WW, 1960 The fallacy of the null-hypothesis significance test. Psychol. Bull 57 (5), 416. [PubMed: 13744252]

Schneider JW, 2015 Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations. Scientometrics 102 (1), 411–432.

Seghier ML, Price CJ, 2018 Interpreting and Utilising Intersubject Variability in Brain Function. Trends Cogn. Sci 22 (6), 517–530. 10.1016/j.tics.2018.03.003. [PubMed: 29609894]

Sitzer DI, Twamley EW, Jeste DV, 2006 Cognitive training in Alzheimer's disease: a meta-analysis of the literature. Acta Psychiatr. Scand 114 (2), 75–90. [PubMed: 16836595]

Stein DG, 2015 Embracing failure: What the Phase III progesterone studies can teach about TBI clinical trials. Brain Inj 29 (11), 1259–1272 Epub 2015 Aug 14. Review. [PubMed: 26274493]

Warren M, 2018 First analysis of 'pre-registered' studies shows sharp rise in null findings 10 10.1038/d41586-018-07118-1.

Wasserstein RL, Schirm AL, & Lazar NA (2019) Moving to a world beyond "p < 0. 05". The American Statistician, 73(sup1), 1–19.

Wykes T, Huddy V, Cellard C, McGurk SR, Czobor P, 2011 A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. Am. J. Psychiatr 168 (5), 472–485. [PubMed: 21406461]