SCIENCE FORUM

How failure to falsify in high-volume science contributes to the replication crisis

Abstract The number of scientific papers published every year continues to increase, but scientific knowledge is not progressing at the same rate. Here we argue that a greater emphasis on falsification – the direct testing of strong hypotheses – would lead to faster progress by allowing well-specified hypotheses to be eliminated. We describe an example from neuroscience where there has been little work to directly test two prominent but incompatible hypotheses related to traumatic brain injury. Based on this example, we discuss how building strong hypotheses and then setting out to falsify them can bring greater precision to the clinical neurosciences, and argue that this approach could be beneficial to all areas of science.

SARAH M RAJTMAJER, TIMOTHY M ERRINGTON AND FRANK G HILLARY*

Background and motivation

The “replication crisis” in various areas of research has been widely discussed in journals over the past decade [see, for example, Gilbert et al., 2016; Baker, 2016; Open Science Collaboration, 2015; Munafò et al., 2017]. At the center of this crisis is the concern that any given scientific result may not be reliable; in this way, the crisis is ultimately a question about the collective confidence we have in our methods and results (Alipourfard et al., 2012). The past decade has also witnessed many advances in data science, and “big data” has both contributed to concerns about scientific reliability (Bollier and Firestone, 2010; Calude and Longo, 2017) and also offered the possibility of improving reliability in some fields (Rodgers and Shrout, 2018).

In this article we discuss scientific progress in the clinical neurosciences, and focus on an example related to traumatic brain injury (TBI). Using this example, we argue that the rapid pace of work in this field, coupled with a failure to directly test and eliminate (falsify) hypotheses, has resulted in an expansive literature that lacks the precision necessary to advance science. Instead, we suggest that falsification – where one develops a strong hypothesis, along with methods that can test and refute this hypothesis – should be used more widely by researchers.

Hyperconnectivity hypothesis in brain connectomics

To provide a specific example for the concerns outlined in this critique, we draw from the literature using resting-state fMRI methods and network analysis (typically graph theory, see Caeyenberghs et al., 2017 to examine systems-level plasticity in TBI). Beginning with one of the first papers combining functional neuroimaging and graph theory to examine network topology (Nakamura et al., 2009), an early observation in the study of TBI was that physical disruption of pathways due to focal and diffuse injury results in regional expansion (increase) in strength or number of functional connections. This initial finding was observed in a small longitudinal sample, but then similar effects were observed in other samples (Mayer et al., 2011; Bharath et al., 2015; Hillary et al., 2015; Johnson et al., 2012; Sharp et al., 2011; Iraji et al., 2016) and animal models of TBI (Harris et al., 2016). These findings were summarized in an expansive literature that lacks the precision necessary to advance science.
Table 1. Examples of hypotheses of different strength.

| Type of research/hypothesis | Example |
|-----------------------------|---------|
| Exploratory                 | “We examine the neural correlates of cognitive deficit after brain injury implementing graph theoretical measures of whole brain neural networks” |
| Testable Association        | “We hypothesize that graph theoretical measures of whole brain neural networks predict cognitive deficit after brain injury” |
| Testable/Falsifiable Position (offers possible mechanism and direction/magnitude of expected finding) | “We hypothesize that memory deficits during the first 6 months post injury are due to white matter connection loss and maintain a linear and positive relationship with increased global network path length” |
| Testable/Falsifiable with Alternative Finding (indicates how the hypothesis would and would not be supported) | “We hypothesize that memory deficits during the first 6 months post injury are due to white matter connection loss and maintain a linear and positive relationship with increased global network path length. Diminished global path length in individuals with greatest memory impairment would challenge this hypothesis” |

Both cannot be true as they have been proposed, so which is correct? Even with two apparently contradictory hypotheses in place, there has been no direct testing of these positions against one another to determine the scenarios where either may have merit. Instead, each of these hypotheses remained unconditionally intact and served to support distinct sets of outcomes.

The most important point to be made from this example is not that competing theories in this literature exist. To the contrary, having competing theories for understanding a phenomenon places science in a strong position; the theories can be tested against one another to qualify (or even eliminate) one position. The point is that there have been no attempts to falsify either a hyperconnectivity or disconnection hypothesis, allowing researchers to evoke one or the other depending upon the finding for a given dataset (i.e., disconnection due to white matter loss, or functional “compensation” in the case of hyperconnectivity). What has contributed to this problem is that increasingly complex computational modeling also increases the investigator degrees of freedom, both implicitly and explicitly, to support their hypotheses. In the case of the current example of neural networks, these include selection from a number of brain atlases or other methods for brain parcellation and likewise numerous approaches to neural network definition (see Hallquist and Hillary, 2019). Figure 1 provides a schematic representation of two distinct and simultaneously supported hypotheses in head injury.

To be clear, the approach taken by investigators in this TBI literature is consistent with a
research agenda designed to meet the demands for high publication throughput (more on this below). Examiners publish preliminary findings but remain appropriately tentative in their conclusions given that the sample is small and unexplained factors are numerous. Indeed, a common refrain in many publications is the “need for replication in a larger sample”. As opposed to pre-registering and testing strong hypotheses, investigators are reinforced to identify significant results (any result) for publication. In brain injury work examining network plasticity, investigators have often made general claims that brain injury results in “different” or “altered” connectivity (a problem dating back to early fMRI studies in TBI; Hillary, 2008). While unintentional, imprecise hypotheses increase the likelihood that chance findings are published. The primary consequence is that all findings are “winners”, permitting growing support for either position without movement toward resolution.

Overall, the TBI connectomics literature presents a clear example of a failure to falsify and we argue that it is attributable, at least in part, to the publication of large numbers of papers reporting the results of studies in which small samples were used to examine under-specified hypotheses. This “science-by-volume” approach is exacerbated by the overuse of inappropriate statistical tests, which increases the probability that spurious findings will be reported as meaningful (Button et al., 2013).

The challenges outlined here, where there is a general failure to test and refute strong hypotheses, are not isolated to the TBI literature. Similar issues have been expressed in preclinical studies of stroke (Corbett et al., 2017) in the translational neurosciences where investigators maintain flexible theory and predictions to fit methodological limitations (Macleod et al., 2014; Pound and Ritskes-Hoitinga, 2018; Henderson et al., 2013), and in cancer research where only portions of published data sets provide support for hypotheses (Begley and Ellis, 2012). These factors have likely contributed to the repeated failure of clinical trials to move from animal models to successful Phase III interventions in clinical neuroscience (Tolchin et al., 2020). This example in the neurosciences also mirrors the longstanding problems of co-existing yet inconsistent theories in other disciplines like social psychology (see Watts, 2017).

Figure 1. Two competing theories for functional network response after brain injury. Panel A represents the typical pattern of resting connectivity for the default mode network (DMN) and the yellow box shows a magnified area of neuronal bodies and their axonal projections. Panel B reveals three active neuronal projections (red) that are then disrupted by hemorrhagic lesion of white matter (Panel C). In response to this injury, a hyperconnectivity response (Panel D, left) shows increased signaling to adjacent areas resulting in a pronounced DMN response (Panel D, right). By contrast a disconnection hypothesis maintains that signaling from the original neuronal assemblies is diminished due to axonal degradation and neuronal atrophy secondary to cerebral diaschisis (Panel E, left) resulting in reduced functional DMN response (Panel E, right).
Big data and computational methods as friend and foe

The big data revolution and advancement of computational modeling powered by enhanced computing infrastructure, on the one hand, has magnified concerns about scientific reliability through unprecedented flexibility in data exploration and analysis. Sufficiency large datasets provably contain spurious correlations and the number of these coincidental regularities increases as the dataset size increases (Calude and Longo, 2017; Graham and Spencer, 1990). Adding flexibility, predictive algorithms built on top of these large datasets typically involve a great number of investigator decisions – the combined effects of which undermine reliability of findings [for an example in connectivity modeling see Hallquist and Hillary, 2019]. Results of
machine learning models, for example, are sensitive to model specification and parameter tuning (Pineau, 2021; Bouthillier et al., 2019; Cwiek et al., 2021). Computational approaches permit systematically combing through a great number of potential variables of interest and their statistical relationships (specifically, at scales which would be manually infeasible). Consequently, the burden of reliability falls upon the existence of strong, well-founded hypotheses with sufficient power and clear pre-analysis plans. It has even been suggested that null hypothesis significance testing should only be used in the neurosciences in support of pre-registered hypotheses based on strong theory (Szucs and Ioannidis, 2017).

So, while there is concern that Big Data moves too fast and without the necessary constraints of theory, there is also emerging sentiment that the tremendous computational power coupled with unparalleled data access has the potential to transform some of the most basic scientific tenets, including introduction of a “third scientific pillar” to be added to theory and experimentation (see National Science Foundation, 2010). While this latest position received criticism (Andrews, 2012), computational methods have been reliably demonstrated to offer novel tools to address the replication crisis – an issue addressed in greater detail in “solutions” below.

Operating without anchors in a sea of high-volume science
One challenge then is to determine where the bedrock of our field (our foundational knowledge) ends, and where areas of discovery that show promise (but have yet to be established) begin. By some measure neurology is a fledgling field in the biological sciences: the publication of De humani corporis fabrica by Vesalius in 1543 is often taken to mark the start of the study of human anatomy (Vesalius, 1555). Jean-Martin Charcot – often referred to as the “founder of neurology” – arrived approximately 300 years later (Zalc, 2018). If we simplify our task and start with the work of Milner, Geschwind and Luria in the 1950s, it is still a challenge to determine what is definitively known and what remains conjectural in the field. This challenge is amplified by the pressure on researchers to publish or perish (Macleod et al., 2014; Kiai, 2019; Lindner et al., 2018). The number of papers published per year continues to increase without asymptote (Bornmann and Mutz, 2015). When considering all papers published in the clinical neurosciences since 1900, more than 50% of the entire literature has been published in the last 10 years and 35% in the last five years (see supplementary figures S1a,b in Priestley et al., 2022). In the most extreme examples, “hyperprolific” lab directors publish a scientific paper roughly every 5 days (Ioannidis et al., 2018). It is legitimate to ask if the current proliferation of published findings has been matched by advances in scientific knowledge, or if the rate of publishing is outpacing scientific ingenuity (Sandström and van den Besselaar, 2016) and impeding the emergence of new theories (Chu and Evans, 2021).

We argue that a culture of science-by-volume is problematic for the reliability of science, primarily when paired with research agendas not designed to test/refute hypotheses. First, without pruning possible explanations through falsification, the science-by-volume approach creates an ever-expanding search space where finite human and financial resources are deployed to maximize breadth in published findings as opposed to depth of understanding (Figure 2A). Second, and as an extension of the last point, failure to falsify in a high-volume environment challenges our capacity to know which hypotheses represent foundational theory, which hypotheses are encouraging but require further confirmation, and which hypotheses should be rejected. Finally, in the case of the least-publishable-unit (Broad, 1981) a single data set may be carved into several smaller papers resulting in circles of self-citation and the illusion of reliable support for a hypothesis (or hypotheses) (Gleeson and Biddle, 2000).

There have even been efforts internationally to make science more deliberate through de-emphasis of publication rates in academic circles (Dijkstra and Bilsen, 2013). Executing this type of systemic change in publication rate poses significant challenges and may ultimately be counter-productive because it fails to acknowledge the advancements in data aggregation and analysis afforded by high performance computing and rapid scientific communication through technology. So, while an argument can be made that our rate of publishing is not commensurate with our scientific progress, a path backward to a lower annual publication rate seems an unlikely solution and ignores the advantages of modernity. Instead, we should work toward establishing scientific foundation by testing and refuting strong hypotheses and these efforts may hold the greatest benefit when used to prune theories to determine the fittest prior to replication (Figure 2B). This effort maximizes resources and makes the goals for replication, as a confrontation of theoretical expectations, very clear (Nosek and Errington, 2020a). The remainder of the paper...
outlines how this can be achieved with focus on several contributors to the replication crisis.

**Accelerating science by falsifying strong hypotheses**

**In praise of strong hypotheses**

Successful refutation of hypotheses ultimately depends upon a number of factors, not the least of which is the specificity of the hypothesis (*Earp and Trafimow, 2015*). A simple, but well-specified, hypothesis, brings greater leverage to science than a hypothesis that is far reaching with broad implications but cannot be directly tested or refuted. Even Popper wrote about concerns in the behavioral sciences regarding the rather general nature of hypotheses (*Bartley, 1978*), a sentiment that has recently been described as a “crisis” in psychological theory advancement (*Rzhetsky et al., 2015*). As discussed in the TBI connectomics example, hypotheses may have been broad and "exploratory" because authors remained conservative in their claims and conclusions because studies have been systematically under-powered (one report estimating power at 8%; *Button et al., 2013*). While exploration is a vital part of science (*Figure 2*), it must be recognized as scientific exploration as opposed to an empirical test of a hypothesis. Under-developed hypotheses have been argued to be at least one contributor to repeated failure of clinical trials in acute neurological interventions (*Schwamm, 2014*) yet, paradoxically, strong hypotheses may offer increased sensitivity to subtle effects even in small samples (*Lazic, 2018*).

If we appeal to Popper, the strongest hypotheses make “risky predictions”, therefore prohibiting alternative explanations (see *Popper, 1963*). Moreover, the strongest hypotheses make clear at the outset the findings that would support the prediction, and also those that would not. Practically speaking this could take the form of teams of scientists developing opposing sets of hypotheses and then agreeing on both the experiments and the outcomes that would falsify one or both positions (what Nosek and Errington refer to as precommitment; *Nosek and Errington, 2020b*). This creates scenarios a priori where strong hypotheses are matched with methods that can provide clear tests. This approach is currently being applied in the “accelerating research on consciousness” programme funded by the Templeton World Charity Foundation. Strong hypotheses must be matched with methods that can provide clear tests, a coupling that cannot be overstated. In the brain imaging literature alone, there are poignant examples where flawed methods (or misunderstanding of their applications) has resulted in the repeated substantiation of spurious results (in structural covariance analysis see *Carmon et al., 2020* in resting-state fMRI see *Satterthwaite et al., 2012; Van Dijk et al., 2012*).

**Addressing heterogeneity to create strong hypotheses**

One approach to strengthen hypotheses is to address sample and methodological heterogeneity which plagues the clinical neurosciences (*Benedict and Zivadinov, 2011; Bennett et al., 2019; Schrag et al., 2019; Zucchella et al., 2020; Yeates et al., 2019*). To echo a recent review of work in the social sciences, the neurosciences require a “heterogeneity revolution” (*Bryan et al., 2021*). Returning again to the TBI connectomics example, researchers relied upon small datasets heterogeneous with respect to age of injury, time post injury, injury severity, and other factors that could critically influence the response of the neural system to injury. Strong hypotheses determine the influence of sample characteristics by directly modeling the effects of demographic and clinical factors (*Bryan et al., 2021*) as opposed to statistically manipulating the variance accounted for by them — including the widespread and longstanding misapplication of covariance statistics to “equilibrate” groups in case-control designs (*Miller and Chapman, 2001; Zinbarg et al., 2010; Storandt and Hudson, 1975*). Finally, strong hypotheses leverage the pace of our current science as an ally, where studies designed specifically to address sample heterogeneity can test the role of clinical and demographic predictors in brain plasticity and outcome.

**Open science and sharing to bolster falsification efforts**

Addressing sample heterogeneity requires large diverse samples, and one way to achieve this is via data sharing. While data-sharing practices and availability differ across scientific disciplines (*Tedersoo et al., 2021*), there are enormous opportunities for sharing data in the clinical neurosciences (see, for example the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) initiative), even in cases where data were not collected with identical methods (such as the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium; see *Olsen et al., 2021* for more on severe brain injury, and *Thompson*).
et al., 2020 for a broad summary of work in clinical neuroscience). However, data aggregation and harmonization approaches remain largely untested as a solution to science-by-volume problems in the neurosciences.

It should be stressed that data sharing as a practice is not a panacea to poor study design and/or an absence of theory. The benefits of data combination do not eliminate any existing issues related to instrumentation and data collection occurring at individual sites; it is crucial to understand that data sharing permits faster accumulation of data while retaining any existing methodological concerns (e.g., harmonization). If unaddressed, these concerns introduce magnified noise or systematic bias masquerading as high-powered findings (Maikusa et al., 2021). However, well-designed data sharing efforts with rigorous harmonization approaches (e.g., Fortin et al., 2017; Tate et al., 2021) hold opportunities for falsification through meta-analyses, mega-analyses, and between site data comparisons (Thompson et al., 2022). Data sharing and team science also provide realistic opportunities to address sample heterogeneity and site-level idiosyncrasies in method.

Returning to the TBI connectomics example above, data sharing could play a central role in resolving this literature. The neural network response to injury most likely depends upon where one looks (specific neural networks), time post injury, and perhaps a range of clinical and demographic factors such as age of injury, current age, sex, and premorbid status. Clinically and demographically heterogeneous samples of n~40–50 subjects do not have the resolution necessary to determine when hyperconnectivity occurs and when it may give way to disconnection (see Caeyenberghs et al., 2017; Hillary and Grafman, 2017). Data sharing and team science organized to test strong hypotheses can provide clarity to this literature.

Harnessing big data to advance metascience

Metascience (Peterson and Panofsky, 2014) has become central to many of the issues raised here. Metascience uses the tools of science to describe and evaluate science on a macro scale and to motivate reforms in scientific practice (Munafò et al., 2017; Ioannidis et al., 2015; Gurevitch et al., 2018). The emergence of metascience is at least partially attributable to advances in web search and indexing, network science, natural language processing, and computational modeling. Amongst other aims, work under this umbrella has sought to diagnose biases in research practice (Larivière et al., 2013; Clauset et al., 2015; Huang et al., 2020), understand how researchers select new work to pursue (Rzhetsky et al., 2015; Jia et al., 2020), identify factors contributing to academic productivity (Liu et al., 2018; Li et al., 2018; Pluchino et al., 2019; Janosov et al., 2020), and forecast the emergence of new areas of research (Prabhakaran et al., 1959; Asoja et al., 2016; Salatino et al., 2018; Chen et al., 2017; Krenn and Zeilinger, 2020; Behrouzi et al., 2020).

A newer thread of ongoing efforts within the metascience community is working to build and promote infrastructure for reproducible and transparent scholarly communication (see Konkol et al., 2020 for a recent review, Wilkinson et al., 2016; Nosek et al., 2015). As part of this vision, primary deliverables of research processes include machine-readable outputs that can be queried by researchers for meta-analyses and theory development (Priem, 2013; Lakens and DeBruine, 2021; Brinckman et al., 2019). These efforts are coupled with recent major investments in approaches to further automate research synthesis and hypothesis generation. The Big Mechanism program, for example, was set up by the Defense Advanced Research Projects Agency (DARPA) to fund the development of technologies to read the cancer biology literature, extract fragments of causal mechanisms from publications, assemble these mechanisms into executable models, and use these models to explain and predict new findings, and even test these predictions (Cohen, 2015).

Lines of research have also emerged using creative assembly of experts (e.g., prediction markets; Dreber et al., 2015; Camerer et al., 2016; Camerer et al., 2018; Gordon et al., 2020 and AI-driven approaches Altmejd et al., 2019; Pawel and Held, 2020; Yang et al., 2020) to estimate confidence in specific research hypotheses and findings. These too have been facilitated by advances in information extraction, natural language processing, machine learning, and larger training datasets. The DARPA-funded Systematizing Confidence in Open Research and Evidence (SCORE) program, for example, is nearing the end of coordinated three-year long effort to develop technologies to predict and explain replicability, generalizability and robustness of published claims in the social and behavioral sciences literatures (Allipourfard et al., 2012). As it continues to advance, the metascience community may serve to revolutionize
the research process resulting in a literature that is readily interrogated and upon which strong hypotheses can be built.

**Falsification for scaffolding convergence research**

Advances in computing hold the promise of richer datasets, AI-driven meta-analyses, and even automated hypothesis generation. However, thus far, efforts to harness big data and emerging technologies for falsification and replication have been relatively uncoordinated, with the aforementioned Big Mechanism and SCORE programs amongst a few notable exceptions.

The need to refine theories becomes increasingly apparent when confronted with resource, ethical, and practical constraints that limit what can be further pursued empirically. At the same time, addressing pressing societal needs requires innovation and convergence research. An example are calls for “grand challenges”, a family of initiatives focused on tackling daunting unsolved problems with large investments intended to make an applied impact. These targeted investments tend to lead to a proliferation of science; however, these mechanisms could also incorporate processes to refine and interrogate theories as they progress towards addressing a specific and compelling issue. A benefit of incorporating falsification into this pipeline is that it encourages differing points of view, a desired feature of grand challenges (Helbing, 2012) and other translational research programs. For example, including clinical researchers in the design of experiments being conducted at the preclinical stage can strengthen the quality of hypotheses before testing them to potentially increase the utility of the result, regardless of the outcome (Seyhan, 2019). To realize the full potential, investment in developing and maturing computational models is also needed to leverage the sea of scientific data to help identify the level of confidence in the fitness and replicability of each theory, and where best to deploy resources. This could lead to more rapid theory refinements and greater feedback for what new data to collect than would be possible using hypothesis-driven or data-intensive approaches in isolation (Peters et al., 2014).

**Practical challenges to falsification**

We have proposed that falsification of strong hypothesis provides a mechanism to increase study reliability. High volume science should ideally function to eliminate possible explanations, otherwise productivity obfuscates progress. But can falsification ultimately achieve this goal? A strict Popperian approach, that every observation represents either a confirmation or refutation of a hypothesis, is challenging to implement in day-to-day scientific practice (Lakatos, 1970; Kuhn, 1970). What’s more, one cannot, with complete certainty, disprove a hypothesis any more than one can hope to prove a hypothesis (see Lakatos, 1970). It was Popper who emphasized that truth is ephemeral and even when it can be accessed, it remains provisional (Popper, 1959).

The philosophical dilemma in establishing the “true” nature of a scientific finding is reflected in the pragmatic challenges facing replication science. Even after an effort to replicate a finding, when investigators are presented with the results and asked if the replication was a success, the outcome is often disagreement resulting in “intellectual gridlock” (Nosek and Errington, 2020b). So, if the goal to falsify a hypothesis is both practically and philosophically flawed, why the emphasis? The answer is that, while falsification cannot remove the foibles of human nature, systematic methodological error, and noise from the scientific process, by setting our sights on testing and refuting strong a priori hypotheses we may uncover the shortcomings to our explanations. Attempts to falsify through refutation cannot be definitive but the outcome of multiple efforts can critically inform the direction of a science (Earp and Trafimow, 2015) when formally integrated into the scientific process (as depicted in Figure 2).

Finally, falsification alone serves as an incomplete response to problems of scientific reliability but becomes a powerful tool when combined with efforts that maximize transparency in method, make null results available, facilitate data/code sharing, and increase the incentive structures for investigators to refocus on open and transparent science.

**Conclusion**

Due to several factors including a high-volume science culture and previously unavailable computational resources, the empirical sciences have never been more productive. This unparalleled productivity invites questions about the rigor and direction of science and, ultimately, how these efforts translate to scientific advancement. We have proposed that it should be a primary goal to identify the “ground truths” that can serve as a foundation for more deliberate study and, to do so, there must be greater emphasis on testing and refuting strong hypotheses. The falsification of strong hypotheses enhances the power of replication first by pruning options...
and identifying the most promising hypotheses including possible mechanisms. When conducted through a team science framework, the endeavor leverages shared datasets that allow us to address heterogeneity that makes so many findings tentative. We must take steps toward more transparent and open science including – and most importantly – study pre-registration of strong hypotheses. The ultimate goal is to harness the rapid advancements in big data, computational power, and strong, well-defined theory with the goal to accelerate science.

Sarah M Rajtmajer is in the College of Information Sciences and Technology, The Pennsylvania State University, University Park, United States

http://orcid.org/0000-0002-1464-0848

Timothy M Errington is at the Center for Open Science, Charlottesville, United States

http://orcid.org/0000-0002-4959-5143

Frank G Hillary is in the Department of Psychology and the Social Life and Engineering Sciences Imaging Center, The Pennsylvania State University, University Park, United States

fhillary@psu.edu

http://orcid.org/0000-0002-1427-0218

Author contributions: Sarah M Rajtmajer, Writing – original draft, Writing – review and editing; Timothy M Errington, Writing – review and editing; Frank G Hillary, Conceptualization, Writing – original draft, Writing – review and editing

Competing interests: The authors declare that no competing interests exist.

Received 22 March 2022
Accepted 28 July 2022
Published 08 August 2022

Funding

No external funding was received for this work.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.78830.sa1
Author response https://doi.org/10.7554/eLife.78830.sa2

Data availability

There are no data associated with this article.

References

Alipourfard N, Arendt B, Benjamin DM, Benkler N, Bishop MM, Burstein M, Bush M, Caverlee J, Chen Y, Clark C, Dreber A, Errington TM, Fidler F, Fox NW, Frank A, Fraser H, Friedman S, Gelman B, Gentile J, Giles CL, et al. 2012. Systematizing Confidence in Open Research and Evidence (SCORE). SocArXiv. https://osf.io/preprints/socarxiv/46mm

Altmjeid A, Dreber A, Forsell E, Huber J, Imai T, Johannesson M, Kirchner M, Nave G, Camerer C, Wicherts JM. 2019. Predicting the replicability of social science lab experiments. PLOS ONE 14:e0225826. DOI: https://doi.org/10.1371/journal.pone.0225826, PMID: 31805105

Andrews GE. 2012. Drowning in the data deluge. Notices of the American Mathematical Society 59:933. DOI: https://doi.org/10.1090/noti871

Asooja K, Bordea G, Vulcu G, Buelerar P. 2016. Forecasting emerging trends from scientific literature. Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016.

Baker M. 2016. 1,500 scientists lift the lid on reproducibility. Nature 533:452–454. DOI: https://doi.org/10.1038/533452a, PMID: 27225100

Bartley WW. 1978. The philosophy of Karl Popper. Philosophy 7:675–716. DOI: https://doi.org/10.1007/BF02378843

Begley CG, Ellis LM. 2012. Drug development: Raise standards for preclinical cancer research. Nature 483:531–533. DOI: https://doi.org/10.1038/483531a, PMID: 22460880

Behrouzi S, Shafaeipour Sarmoor Z, Hajsadeghi K, Kavousi K. 2020. Predicting scientific research trends based on link prediction in keyword networks. Journal of Informetrics 14:101079. DOI: https://doi.org/10.1016/j.joi.2020.101079

Benedict RHB, Zivadinov R. 2011. Risk factors for and management of cognitive dysfunction in multiple sclerosis. Nature Reviews Neurology 7:332–342. DOI: https://doi.org/10.1038/nrneurol.2011.61, PMID: 21556031

Bennett SD, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, Manzotti G, Shafran R. 2019. Practitioner review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: A systematic review and meta-analysis. Journal of Child Psychology and Psychiatry, and Allied Disciplines 60:828–847. DOI: https://doi.org/10.1111/jcpp.13010, PMID: 30775782

Bharath RD, Munivenkatappa A, Gehel S, Panda R, Saini J, Rajeswaran J, Shukla D, Bhagavatula ID, Biswal BB. 2015. Recovery of resting brain connectivity ensuing mild traumatic brain injury. Frontiers in Human Neuroscience 9:513. DOI: https://doi.org/10.3389/fnhum.2015.00513, PMID: 26441610

Bollier D, Firestone CM. 2010. The promise and peril of big data. Aspen Institute, Communications and Society Program. https://www.aspeninstitute.org/publications/promise-peril-big-data/ [Accessed August 2, 2022].

Bonnelle V, Leech R, Kinnunen KM, Ham TE, Beckmann CF, De Boisseyon X, Greenwood RJ, Sharp DJ. 2011. Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. Journal of Neuroscience 31:13442–13451. DOI: https://doi.org/10.1523/JNEUROSCI.1163-11.2011, PMID: 21940437

Bortolotti L, Mutz R. 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66:2215–2222. DOI: https://doi.org/10.1002/asi.23329

Boulter S, Buitelaar J, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, Manzotti G, Shafran R. 2019. Practitioner review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: A systematic review and meta-analysis. Journal of Child Psychology and Psychiatry, and Allied Disciplines 60:828–847. DOI: https://doi.org/10.1111/jcpp.13010, PMID: 30775782

Bharath RD, Munivenkatappa A, Gehel S, Panda R, Saini J, Rajeswaran J, Shukla D, Bhagavatula ID, Biswal BB. 2015. Recovery of resting brain connectivity ensuing mild traumatic brain injury. Frontiers in Human Neuroscience 9:513. DOI: https://doi.org/10.3389/fnhum.2015.00513, PMID: 26441610

Bollier D, Firestone CM. 2010. The promise and peril of big data. Aspen Institute, Communications and Society Program. https://www.aspeninstitute.org/publications/promise-peril-big-data/ [Accessed August 2, 2022].

Bonnelle V, Leech R, Kinnunen KM, Ham TE, Beckmann CF, De Boisseyon X, Greenwood RJ, Sharp DJ. 2011. Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. Journal of Neuroscience 31:13442–13451. DOI: https://doi.org/10.1523/JNEUROSCI.1163-11.2011, PMID: 21940437

Bortolotti L, Mutz R. 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66:2215–2222. DOI: https://doi.org/10.1002/asi.23329

Boulter S, Buitelaar J, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, Manzotti G, Shafran R. 2019. Practitioner review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: A systematic review and meta-analysis. Journal of Child Psychology and Psychiatry, and Allied Disciplines 60:828–847. DOI: https://doi.org/10.1111/jcpp.13010, PMID: 30775782

Bharath RD, Munivenkatappa A, Gehel S, Panda R, Saini J, Rajeswaran J, Shukla D, Bhagavatula ID, Biswal BB. 2015. Recovery of resting brain connectivity ensuing mild traumatic brain injury. Frontiers in Human Neuroscience 9:513. DOI: https://doi.org/10.3389/fnhum.2015.00513, PMID: 26441610

Bollier D, Firestone CM. 2010. The promise and peril of big data. Aspen Institute, Communications and Society Program. https://www.aspeninstitute.org/publications/promise-peril-big-data/ [Accessed August 2, 2022].

Bonnelle V, Leech R, Kinnunen KM, Ham TE, Beckmann CF, De Boisseyon X, Greenwood RJ, Sharp DJ. 2011. Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. Journal of Neuroscience 31:13442–13451. DOI: https://doi.org/10.1523/JNEUROSCI.1163-11.2011, PMID: 21940437

Bortolotti L, Mutz R. 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66:2215–2222. DOI: https://doi.org/10.1002/asi.23329
reproducibility: capturing the “whole tale.” Future Generation Computer Systems 94:854–867. DOI: https://doi.org/10.1016/j.future.2017.12.029
Broad WJ. 1981. The publishing game: Getting more for less. Science 211:1137–1139. DOI: https://doi.org/10.1126/science.7008199, PMID: 7008199
Bryan CJ, Tipton E, Yeager DS. 2021. Behavioural science is unlikely to change the world without a heterogeneity revolution. Nature Human Behaviour 5:980–989. DOI: https://doi.org/10.1038/s41562-021-01143-3, PMID: 34299401
Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafo MR. 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience 14:365–376. DOI: https://doi.org/10.1038/nrn3475, PMID: 23571845
Caeyenberghs K, Verhelst H, Clemente A, Wilson PH. 2017. Mapping the functional connectome in traumatic brain injury: What can graph metrics tell us? NeuroImage 160:113–123. DOI: https://doi.org/10.1016/j.neuroimage.2016.12.003, PMID: 27919750
Calude CS, Longo G. 2017. The deluge of spurious correlations in big data. Foundations of Science 22:595–612. DOI: https://doi.org/10.1007/s10699-016-9498-4
Camerer CF, Dreber A, Forsell E, Gampa A, Heikensten E, Hummer L, Imai T, Isaksson S, et al. 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour 2:637–644. DOI: https://doi.org/10.1038/s41562-018-0399-z, PMID: 31346273
Carmon J, Heege J, Necus JH, Owen TW, Pipa G, Razen M, Wu H. 2016. Evaluating reliability of laboratory experiments in economics. Science 351:1433–1436. DOI: https://doi.org/10.1126/science.aaf0918, PMID: 32874648
Chen C, Wang Z, Li W, Sun X. 2017. Modeling scientific influence for research trending topic prediction. Proceedings of the AAAI Conference on Artificial Intelligence. . DOI: https://doi.org/10.1609/aaai.v32i1.11882
Chu JSG, Evans JA. 2021. Slowed canonical progress in large fields of science. PNAS 118:e2021636118. DOI: https://doi.org/10.1073/pnas.2021636118, PMID: 34607941
Clauset A, Arbesman S, Larrewore DB. 2015. Systematic inequality and hierarchy in faculty hiring networks. Science Advances 1:e1400005. DOI: https://doi.org/10.1126/sciadv.1400005, PMID: 26601125
Cohen PR. 2015. DARPA’s Big Mechanism program. Physical Biology 12:045008. DOI: https://doi.org/10.1088/1478-3975/12/4/045008, PMID: 26178259
Corbett D, Carmichael ST, Murphy TH, Jones TA, Schwab ME, Jolkonen J, Clarkson AN, Dancause N, Weiloch T, Johansen-Berg H, Nilsson M, McCullough LD, Joy MT. 2017. Enhancing the alignment of the preclinical and clinical stroke recovery research pipeline: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable Translational Working Group. Neurorehabilitation and Neural Repair 31:699–707. DOI: https://doi.org/10.1177/1545963817724285, PMID: 28803530
Cwiek A, Rajtmajer SM, Wyble B, Honavar V, Grossner E, Hillary FG. 2021. Feeding the machine: challenges to reproducible predictive modeling in resting-state connectomics. Network Neuroscience 1:1–20. DOI: https://doi.org/10.1162/netn_a_00212, PMID: 35350584
Dijkstra bloem H, Miedema F, Huisman F, Mijnhardt W. 2013. Position paper: Why science does not work as it should and what to do about it. http://scienceintransition.nl/app/uploads/2013/10/Science-in-Transition-Position-Paper-final.pdf [Accessed August 2, 2022], Dreber A, Pfeiffer T, Almenberg J, Isaksson S, Wilson B, Chen Y, Nosek BA, Johannesson M, Wachter KW. 2015. Using prediction markets to estimate the reproducibility of scientific research. PNAS 112:15343–15347. DOI: https://doi.org/10.1073/pnas.1516179112, PMID: 26553988
Earp BD, Trafimow D. 2015. Replication, falsification, and the crisis of confidence in social psychology. Frontiers in Psychology 6:621. DOI: https://doi.org/10.3389/fpsyg.2015.00621, PMID: 26042061
Fagerholm ED, Hellyer PJ, Scott G, Leech R, Sharp DJ. 2015. Disconnection of network hubs and cognitive impairment after traumatic brain injury. Brain: A Journal of Neurology 138:1696–1709. DOI: https://doi.org/10.1093/brain/awv075, PMID: 25808370
Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE, Schultz RT, Verma R, Shinozaka RT. 2017. Harmonization of multi-site diffusion tensor imaging data. NeuroImage 161:149–170. DOI: https://doi.org/10.1016/j.neuroimage.2017.08.047, PMID: 28826946
Gilbert DT, King G, Pettigrew S, Wilson TD. 2016. Replication, falsification, and the crisis of confidence in social psychology. Science 351:1037. DOI: https://doi.org/10.1126/science.aad7243, PMID: 26941311
Gleeson M, Biddle S. 2000. Duplicate publishing and the least publishable unit. Journal of Sports Sciences 18:227–228. DOI: https://doi.org/10.1080/026404100364956, PMID: 10824638
Gordon M, Viganola D, Bishop M, Chen Y, Dreber A, Goldfedder B, Holzmeister F, Johannesson M, Liu Y, Trafimow D. 2015. Replication, falsification, and the crisis of confidence in social psychology. Science 351:1037. DOI: https://doi.org/10.1126/science.aad7243, PMID: 26941311
Grossner E, Hillary FG. 2021. Feeding the machine: challenges to reproducible predictive modeling in resting-state connectomics. Network Neuroscience 1:1–20. DOI: https://doi.org/10.1162/netn_a_00212, PMID: 35350584
Harmonization of multi-site diffusion tensor imaging data. NeuroImage 161:149–170. DOI: https://doi.org/10.1016/j.neuroimage.2017.08.047, PMID: 28826946
Gilbert DT, King G, Pettigrew S, Wilson TD. 2016. Comment on “Estimating the reproducibility of psychological science.” Science 351:1037. DOI: https://doi.org/10.1126/science.aad7243, PMID: 26941311
Gleeson M, Biddle S. 2000. Duplicate publishing and the least publishable unit. Journal of Sports Sciences 18:227–228. DOI: https://doi.org/10.1080/026404100364956, PMID: 10824638
Gordon M, Viganola D, Bishop M, Chen Y, Dreber A, Goldfedder B, Holzmeister F, Johannesson M, Liu Y, Trafimow D, Wang J, Pfeiffer T. 2020. Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. Royal Society Open Science 7:200566. DOI: https://doi.org/10.1098/rsos.200566, PMID: 32874648
Graham RL, Spencer JH. 1990. Ramsey theory. Scientific American 263:112–117. DOI: https://doi.org/10.1038/scientificamerican0790-112
Gurevitch J, Koricheva J, Nakagawa S, Stewart G. 2018. Meta-analysis and the science of research synthesis. Nature 555:175–182. DOI: https://doi.org/10.1038/nature25753, PMID: 29517004
Hallquist MN, Hillary FG. 2019. Graph theory approaches to functional network organization in brain networks. Science Forum | How failure to falsify in high-volume science contributes to the replication crisis
disorders: A critique of brave new small-world. Network Neuroscience 3:1–26. DOI: https://doi.org/10.1162/netn_a_00054, PMID: 30793071

Harris NG, Verley DR, Gutman BA, Thompson PM, Yeh HJ, Brown JA. 2016. Disconnection and hyperconnectivity underlie reorganization after TBI: A rodent functional connectomic analysis. Experimental Neurology 277:124–138. DOI: https://doi.org/10.1016/j.expneurol.2015.12.020, PMID: 26730520

Helsing D. 2012. Accelerating scientific discovery by formulating grand scientific challenges. The European Physical Journal Special Topics 214:41–48. DOI: https://doi.org/10.1140/epjst/e2012-01687-x

Henderson VC, Kimmelman J, Ferguson D, Grimshaw JM, Hackam DG. 2013. Threats to validity in the design and conduct of preclinical efficacy studies: A systematic review of guidelines for in vivo animal experiments. PLOS Medicine 10:e1001489. DOI: https://doi.org/10.1371/journal.pmed.1001489, PMID: 23935460

Hillery FG. 2008. Neuroimaging of working memory dysfunction and the dilemma with brain reorganization hypotheses. Journal of the International Neuropsychological Society 14:526–534. DOI: https://doi.org/10.1017/S1355617708080788, PMID: 18577281

Hillery FG, Roman CA, Venkatesan U, Rajtmajer SM, Bajo R, Castellanos ND. 2015. Hyperconnectivity is a fundamental response to neurological disruption. Neuropsychology 29:59–75. DOI: https://doi.org/10.1037/neu0000110, PMID: 24933491

Hillery FG, Graffam JH. 2017. Injured brains and adaptive networks: the benefits and costs of hyperconnectivity. Trends in Cognitive Sciences 21:385–401. DOI: https://doi.org/10.1016/j.tics.2017.03.003, PMID: 28372878

Huang J, Gates AJ, Sinatra R, Barabási AL. 2020. Success and luck in creative careers. EPJ Data Science 9:1–13. DOI: https://doi.org/10.1038/s41562-017-0078

Johnson B, Zhang K, Gay M, Horovitz S, Hallett M, Sebastianelli W, Slobounov S. 2012. Alteration of brain default network in subacute phase of injury in concussed individuals: Resting-state fMRI study. NeuroImage 59:511–518. DOI: https://doi.org/10.1016/j.neuroimage.2011.07.081, PMID: 21846504

Kia A. 2019. To protect credibility in science, banish “publish or perish.” Nature Human Behaviour 3:1017–1018. DOI: https://doi.org/10.1038/s41562-019-0741-0, PMID: 31602039

Konkol M, Nüst D, Goulieri. 2020. Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication. Research Integrity and Peer Review 5:10. DOI: https://doi.org/10.1186/s41073-020-00095-y, PMID: 32685199

Krenn M, Zeilinger A. 2020. Predicting research trends with semantic and neural networks with an application in quantum physics. PNAS 117:1910–1916. DOI: https://doi.org/10.1073/pnas.1914370116, PMID: 31937664

Kuhn TS. 1970. Logic of discovery or psychology of research. Lakatos I, Musgrave A (Eds). Criticism and the Growth of Knowledge. Cambridge: Cambridge University Press. p. 1–23. DOI: https://doi.org/10.1017/CBO9781139171434

Lakates I. 1970. History of science and its rational reconstructions. PSA 1970:91–136. DOI: https://doi.org/10.1080/psaprobnet.1970.495757

Lakens D, DeBruine LM. 2021. Improving transparency, falsifiability, and rigor by making hypothesis tests machine-readable. Advances in Methods and Practices in Psychological Science 4:251524592097094. DOI: https://doi.org/10.1177/2515245920970949

Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. 2013. Bibliometrics: Global gender disparities in science. Nature 504:211–213. DOI: https://doi.org/10.1038/504211a, PMID: 24350369

Lazic SE. 2018. Four simple ways to increase power without increasing the sample size. Laboratory Animals 52:621–629. DOI: https://doi.org/10.1177/0023677218764748, PMID: 29629616

Li W, Aste T, Caccioli F, Livan G. 2018. Early coauthorship with top scientists predicts success in academic careers. Nature Communications 10:5170. DOI: https://doi.org/10.1038/s41467-019-13130-4, PMID: 31729362

Lindner MD, Torralba KD, Khan NA, Ouzounis CA. 2018. Scientific productivity: An exploratory study of metrics and incentives. PLOS ONE 13:e0195321. DOI: https://doi.org/10.1371/journal.pone.0195321, PMID: 29614101

Liu L, Wang Y, Sinatra R, Giles CL, Song C, Wang D. 2018. Hot streaks in artistic, cultural, and scientific careers. Nature 559:396–399. DOI: https://doi.org/10.1038/s41586-018-0315-8, PMID: 29995850

Macleod MR, Michie S, Roberts I, Dirnagl U, MacLaren FG. 2015. Bibliometrics: Global gender disparities in science. Nature 526:41–48. DOI: https://doi.org/10.1038/ncomms17143, PMID: 27284033

Maikusa N, Zhu Y, Uematsu A, Yamashita A, Saotome K, Okada N, Kasai K, Okanoya K, Yamashita O, Tanaka SC, Koike S. 2021. Comparison of traveling-subject and combat harmonization methods for assessing structural brain characteristics. Human Brain Mapping 1:5278–5287. DOI: https://doi.org/10.1002/hbm.25615
Mayer AR, Mannell MV, Ling J, Gasparovic C, Yeo RA. 2011. Functional connectivity in mild traumatic brain injury. Human Brain Mapping 32:1825–1835. DOI: https://doi.org/10.1002/hbm.21151, PMID: 21259381
Miller GA, Chapman JP. 2001. Misunderstanding analysis of covariance. Journal of Abnormal Psychology 110:40–48. DOI: https://doi.org/10.1037//0021-843x.110.1.40, PMID: 11261398
Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, Simonson UJ, Wagenmakers EJ, Ware JJ, Ioannidis JPA. 2017. A manifesto for reproducible science. Nature Human Behaviour 1:0021. DOI: https://doi.org/10.1038/s41562-016-0021, PMID: 33954258
Nakamura T, Hillary FG, Biswal BB. 2009. Resting network plasticity following brain injury. PLOS ONE 4:e8220. DOI: https://doi.org/10.1371/journal.pone.0008220, PMID: 20011533
National Science Foundation. 2010. Computational and Data-enabled Science and Engineering. http://www.nsf.gov/mps/cds-e/ (Accessed August 2, 2022).
Nosek BA, Alter G, Banks GE, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, da Fava A, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hedges B, Humphreys M, Ishiyama J, et al. 2015. Promoting an open research culture. Science 348:1422–1425. DOI: https://doi.org/10.1126.science.aab2374
Nosek BA, Errington TM. 2020a. What is replication? PLOS Biology 18:e3000691. DOI: https://doi.org/10.1371/journal.pbio.3000691
Nosek BA, Errington TM. 2020b. The best time to argue about what a replication means? Before you do it. Nature 583:518–520. DOI: https://doi.org/10.1038/s41586-020-02142-6, PMID: 32694846
Olsen A, Babikian T, Bigler ED, Caeyenberghs K, Conde V, Dams-O’Connor K, Dobryakova E, Genova H, Grafman J, Häberg AK, Heggland I, Hellström T, Hodges CB, Irinia A, Jha RM, Johnson PK, Koliatsos VE, Levin H, Li LM, Lindsey HM, et al. 2021. Toward a global and reproducible science for brain imaging in neurotrauma: the ENIGMA adult moderate/severe traumatic brain injury working group. Brain Imaging and Behavior 15:275–279. DOI: https://doi.org/10.1007/s11682-020-00313-7, PMID: 32797398
Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. Science 349:aac4716. DOI: https://doi.org/10.1126/science.aac4716, PMID: 26315443
Pawel S, Held L. 2020. Probabilistic forecasting of replication studies. PLOS ONE 15:e0231416. DOI: https://doi.org/10.1371/journal.pone.0231416, PMID: 32320420
Peters DPC, Havstad KM, Cushing J, Tweedie C, Fuentes O, Villanueva-Rosas N. 2014. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere 5:art67. DOI: https://doi.org/10.1890/ES13-00359.1
Peterson D, Panofsky DPA. 2014. Metascience as a Scientific Social Movement. SocArXiv. https://osf.io/preprints/socarxiv/4dsqa/
Pineau J. 2021. Improving reproducibility in machine learning research: a report from the neurips 2019 reproducibility program. Journal of Machine Learning Research 22:1–20.
Pluchino A, Burgio G, Rapisarda A, Biondo AE, Pulvirenti A, Ferro A, Giorgino T. 2019. Exploring the role of interdisciplinarity in physics: success, talent and luck. PLOS ONE 14:e0218793. DOI: https://doi.org/10.1371/journal.pone.0218793, PMID: 31242227
Popper KR. 1959. The Logic of Scientific Discovery. Julius Springer, Hutchinson & Co.
Popper K. 1963. Conjectures and Refutations: The Growth of Scientific Knowledge. Routledge.
Pound P, Ritskes-Hoitinga M. 2018. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. Journal of Translational Medicine 16:304. DOI: https://doi.org/10.1186/s12967-018-1678-1, PMID: 30404629
Prabhakaran V, Hamilton WL, McFarland D, Jurafsky D. 1959. Predicting the Rise and Fall of Scientific Topics from Trends in their Rhetorical Framing. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. . DOI: https://doi.org/10.18653/v1/P16-1111
Priem J. 2013. Beyond the paper. Nature 495:437–440. DOI: https://doi.org/10.1038/495437a, PMID: 23538811
Priestley D, Staph J, Koneru S, Rajtmajer S, Hillary F. 2022. Establishing Ground Truth in the Clinical Neurosciences: If Replication Is the Answer, Then What Are the Questions?. PsyArXiv. https://psyarxiv.com/rb32d/
Rodgers JL, Shrut PE. 2018. Psychology’s replication crisis as scientific opportunity: A précis for policymakers. Policy Insights from the Behavioral and Brain Sciences 5:134–141. DOI: https://doi.org/10.1177/2372732217749254
Rzetkowsky A, Foster JG, Foster IT, Evans JA. 2015. Choosing experiments to accelerate collective discovery. PNAS 112:14569–14574. DOI: https://doi.org/10.1073/pnas.1509757112, PMID: 26554009
Salatino AA, Osborne F, Motta E. 2018. AUGUR: Forecasting the Emergence of New Research Topics. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. . DOI: https://doi.org/10.1145/3197026.3197052
Sandström U, van den Besselaar P. 2016. Quantity and/or quality? The importance of publishing many papers. PLOS ONE 11:e0166149. DOI: https://doi.org/10.1371/journal.pone.0166149, PMID: 27870624
Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur RE. 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. NeuroImage 60:623–632. DOI: https://doi.org/10.1016/j.neuroimage.2011.12.063, PMID: 22233733
Schrag A, Zhelev SS, Hotham S, Merritt RD, Khan K, Graham L. 2019. Heterogeneity in progression of network plasticity following brain injury. Human Brain Mapping 440. DOI: https://doi.org/10.1002/hbm.21151, PMID: 21259381
Schramm V, Tholm P, Hotham S, Merritt RD, Khan K, Graham L. 2019. Heterogeneity in progression of network plasticity following brain injury. Human Brain Mapping 440. DOI: https://doi.org/10.1002/hbm.21151, PMID: 21259381
Schwarz J, Koneru S, Rajtmajer S, Hillary F. 2021. Toward reproducibility in the scientific method: An assessment of reproducibility in neuroimaging research. Science Forum | How failure to falsify in high-volume science contributes to the replication crisis

Rajtmajer et al. eLife 2022;11:e78830. DOI: https://doi.org/10.7554/eLife.78830
Translational Medicine Communications 4:18. DOI: https://doi.org/10.1186/s41231-019-0050-7

Sharp DJ, Beckmann CF, Greenwood R, Kinnunen KM, Bonnelle V, De Boissezon X, Powell JH, Counsell SJ, Patel MC, Leech R. 2011. Default mode network functional and structural connectivity after traumatic brain injury. Brain: A Journal of Neurology 134:2233–2247. DOI: https://doi.org/10.1093/brain/war175, PMID: 21841202

Sharp DJ, Scott G, Leech R. 2014. Network dysfunction after traumatic brain injury. Nature Reviews Neurology 10:156–166. DOI: https://doi.org/10.1038/nrneurol.2014.15, PMID: 24514870

Storandt M, Hudson W. 1975. Misuse of analysis of covariance in aging research and some partial solutions. Experimental Aging Research 1:121–125. DOI: https://doi.org/10.1080/03610737508257953, PMID: 1053198

Szucs D, Ioannidis JPA. 2017. When null hypothesis significance testing is unsuitable for research: A reassessment. Frontiers in Human Neuroscience 11:390. DOI: https://doi.org/10.3389/fnhum.2017.00390, PMID: 28824397

Tate DF, Dennis EL, Adams JT, Adamson MM, Belanger HG, Bigler ED, Bouchard HC, Clark AL, Delano-Wood LM, Disner SG, Egan BF, Egan GF, Kochunov P, Thompson PM, Jahanshad N, Schmaal L, Turner JA, et al. 2020. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease. Brain: A Journal of Neurology 143:1348–1348. DOI: https://doi.org/10.1093/brain/awy287, PMID: 30462211

Yeates KO, Tang K, Barrowman N, Freedman SB, Gravel J, Gagnon I, Sangha G, Boutis K, Beer D, Craig W, Burns E, Fanion KJ, Mikrogianakis A, Barlow K, Dubrovsky AS, Meeuwisse W, Gioia G, Meehan WP, Beauchamp MH, Kamil Y, et al. 2019. Validation and initial validation of clinical phenotypes of children presenting with concussion acutely in the emergency department: Latent class analysis of a multi-center, prospective cohort, observational study. Journal of Neurotrauma 36:1758–1767. DOI: https://doi.org/10.1089/neu.2018.6009, PMID: 30618356

Zalc B. 2018. One hundred and fifty years ago Charcot reported multiple sclerosis as a new neurological disease. Brain: A Journal of Neurology 141:3482–3488. DOI: https://doi.org/10.1093/brain/awy287, PMID: 30462211

Zinbarg RE, Suzuki S, Uliazeik AA, Lewis AR. 2010. Biased parameter estimates and inflated type I error rates in analysis of covariance (and analysis of partial variance) arising from unreliability: Alternatives and remedial strategies. Journal of Abnormal Psychology 119:307–319. DOI: https://doi.org/10.1037/a0017552, PMID: 20455604

Zucchella C, Mantovani E, Federico A, Lugoboni F, Tamburin S. 2020. Non-invasive brain stimulation for gambling disorder: A systematic review. Frontiers in Neuroscience 14:729. DOI: https://doi.org/10.3389/fnins.2020.00729, PMID: 33013280