International Studies in the Philosophy of Science

ISSN: 0269-8595 (Print) 1469-9281 (Online) Journal homepage: www.tandfonline.com/journals/cisp20

Is There a Reproducibility Crisis? On the Need for Evidence-based Approaches

Stephen Pethick, Mark N. Wass & Martin Michaelis

To cite this article: Stephen Pethick, Mark N. Wass & Martin Michaelis (2025) Is There a Reproducibility Crisis? On the Need for Evidence-based Approaches, International Studies in the Philosophy of Science, 38:4, 287-303, DOI: 10.1080/02698595.2025.2538937

To link to this article: https://doi.org/10.1080/02698595.2025.2538937

Published online: 28 Jul 2025.

INTERNATIONAL STUDIES IN THE PHILOSOPHY OF SCIENCE
2025, VOL. 38, NO. 4, 287–303
https://doi.org/10.1080/02698595.2025.2538937

Routledge
Taylor & Francis Group

Is There a Reproducibility Crisis? On the Need for Evidence-based Approaches

Stephen Pethick[a], Mark N. Wass[b] and Martin Michaelis [b,c]

[a]Kent Law School, University of Kent, Canterbury, UK; [b]School of Natural Sciences, University of Kent, Canterbury, UK; [c]Dr Petra Joh Research Institute, Frankfurter Stiftung für krebskranke Kinder, Frankfurt am Main, Germany

ABSTRACT

The 'Sixth Report—Reproducibility and Research Integrity' (UK House of Commons Science, Innovation and Technology Committee 2023. 'Sixth Report—Reproducibility and Research Integrity') ('The Report') recommends measures designed to tackle an alleged 'reproducibility crisis' in scientific research. Our systematic analysis of the content of this report revealed that its findings and recommendations are consistent with the scientific literature, including the acknowledgement that conclusive evidence demonstrating the existence of a 'reproducibility crisis' is lacking. Though conceding that there is currently no way to determine the size of the crisis or whether it even exists, The Report nevertheless proposes actions to tackle the alleged crisis. However, without a quantitative understanding, the efficacy of the proposed measures cannot be verified. Hence, the current approach towards the alleged reproducibility crisis, here exemplified by The Report, does not adhere to the standards that would normally applied to the scientific method. An evidence-based approach requires the establishment of a quantitative understanding of whether data variability in the research literature exceeds technically achievable levels of reproducibility. If it does, the resulting understanding will enable the design of actions, whose success can be monitored. Our findings emphasise that the research environment requires the same level of rigour and scrutiny as the scientific experiments themselves.

ARTICLE HISTORY
Received 20 March 2025
Accepted 22 July 2025

KEYWORDS
Evidence; science; innovation and technology committee; reproducibility; research integrity; scientific method

1. Introduction

A 'reproducibility crisis', i.e. a worrying lack of reproducibility of research findings published in peer-reviewed scientific articles, is discussed across many academic disciplines (Ioannidis 2005; Alberts et al. 2015; Begley and Ioannidis 2015; Benjamin et al. 2018; Catillon 2019; Goldacre, Morton, and DeVito 2019; Laraway et al. 2019; Wass, Ray, and Michaelis 2019; Munafò et al. 2020; Munafò et al. 2020; Diaba-Nuhoho and

CONTACT Martin Michaelis m.michaelis@kent.ac.uk School of Natural Sciences, University of Kent, Canterbury CT2 7NJ, UK

Amponsah-Offeh 2021; Stewart et al. 2021; UK Reproducibility Network Steering Committee 2021; Freese, Rauf, and Voelkel 2022; MacLeod et al. 2022; Oransky 2022; Rajtmajer, Errington, and Hillary 2022; Stewart et al. 2022; Wang and Long 2022; Ball 2023; Butler et al. 2023; Oransky, Marcus, and Abritis 2023; Song et al. 2023; Yang et al. 2023; Oransky and Redman 2024). In response to the concerns about reproducibility, on 10th May 2023, the UK House of Commons Science, Innovation and Technology Committee published its 'Sixth Report—Reproducibility and Research Integrity' (House of Commons Science, Innovation and Technology Committee 2023, henceforth, 'The Report'), focusing particularly on reproducibility in academic research.

The evidence was gathered by a call for written submissions, launched on 23rd July 2021, to address the following issues: '(a) whether there was a 'reproducibility crisis' in science and social science research, its scale and causes, and the most vulnerable research areas; (b) the roles of different stakeholders in addressing such problems, relevant policies and mechanisms needed, and the potential contribution of the UKRI's national committee on research integrity; and (c) the measures required for an open, contestable, and rigorous research environment' (House of Commons Science, Innovation and Technology Committee 2021; House of Commons Science, Innovation and Technology Committee 2022; The Report p. 5/6, paragraph 3). In response to the call, 100 pieces of written evidence were submitted. Moreover, 16 witnesses provided oral evidence. The resulting information was synthesised into a report on the extent of the reproducibility crisis (chapter 2) together with recommended solutions (chapters 3–6, The Report).

In our view, The Report serves as a good example of how academia and politics presently interact, in particular exemplifying the processes underlying policymaking on academic matters. In this context, we here provide a systematic analysis of the evidence provided in The Report with a focus on the evidence base of the measures proposed.

2. The Nature of the Reproducibility Crisis According to The Report

In addition to the written and oral evidence provided by individual scientists, groups of scientists, and research organisations, seminal articles in peer-reviewed scientific journals were considered as evidence for the existence of a 'reproducibility crisis'. These included the article 'Why most published research findings are false' by John Ioannidis (Ioannidis 2005), a survey by the journal Nature, in which >70% of the 1576 respondents reported at least one failure to reproduce an experiment of someone else (Baker 2016), and a Wellcome Trust survey, in which 61% of 1,832 junior researchers reported pressure by their supervisors to reproduce particular results and 13% did not feel comfortable to approach their supervisors about unsuccessful replication attempts (Wellcome Trust 2020).

The combined evaluation of data from Retraction Watch and the Centre for Scientific Integrity (reporting the retraction of in total >1,100 research articles by authors with a UK affiliation) and numbers from the UK Department for Business, Energy, and Industrial Strategy (BEIS) (showing that UK researchers publish >200,000 per year) resulted in the conclusions that 'retractions were relatively uncommon but not insignificant' (UK Department for Business, Energy, and Industrial Strategy 2022; The Report, p. 14, paragraph 24).

Based on the written and oral statements, The Report concluded that 'all research disciplines are affected by the systemic issues that limit reproducibility' (The Report, p. 16, paragraph 32) and that 'there is need for action to address the significant problems

caused by the prevalence of reproducibility problems in the scientific community' (The Report, p. 22, paragraph 50). Moreover, The Report welcomed the 'establishment of the UK Committee on Research Integrity (CORI)' (The Report, p. 20, paragraph 43) but noted: 'It is disappointing that UK CORI's recently published strategy did not mention reproducibility, especially since our inquiry highlighted that this is a major research integrity issue. UK CORI should make sure reproducibility challenges are given due attention and not overlooked in deference to other pressing research integrity issues. UK CORI should develop a sub-committee to focus on reproducibility challenges in research. This sub-committee should establish the relative weight of reproducibility within research integrity concerns, and UK CORI should use this evidence to plan its prioritisation'. (The Report, p. 20/21, paragraph 44)

3. Reproducibility Challenges and Proposed Solutions in The Report

Chapters three to six of The Report discuss 'specific solutions' under the following heads: 'conducting research'; 'communicating research'; 'assessing research'; and 'proposed actions for key actors'. The key actors identified in The Report are the Government, research funders, research institutions and groups, individual researchers, and publishers (The Report).

Chapter three (The Report, 'The Government's role', p. 23/24, paragraphs 51–55) concluded that the government has the responsibility to provide a 'framework'. Chapter four ('Solving reproducibility challenges which emerge as research is conducted'), opens with five paragraphs (The Report, p. 25/26, paragraphs 57–61) that discuss misconduct. Consideration is given to the establishment of a body that investigates misconduct and employs experts 'skilled at data sleuthing and capable of evaluating quality of data and analysis in a range of scientific areas' (The Report, p. 25, paragraph 59). Paragraphs 62–68 (The Report, p. 26–28) then focus on methodology, including common malpractices, transparent reporting, data presentation, and the establishment of guidelines. Since best practice is most likely to be achieved if researchers are guided by appropriate incentives, these considerations result in the conclusion: 'Therefore, the research community, including research institutions and publishers, should work alongside individuals to create an environment where research integrity and reproducibility are championed'. (The Report, p. 28, paragraph 68)

Along similar lines, paragraphs 69–72 proposed 'Promoting a culture of reproducibility', in which reproducibility is valued and supported (The Report, p. 28/29), paragraphs 73–77 requested statistical guidance and expertise (The Report, p. 29/30), paragraphs 78–83 suggested appropriate training of researchers (The Report, p. 31), and paragraphs 84–89 argued for an academic structure and funding system that provides researchers with a structure that enables them to apply best practices to improve reproducibility (The Report, p. 32/33).

Chapter 5 ('Solving reproducibility challenges which emerge in the communication of research', The Report, p. 34–41, paragraphs 90–123) focused on issues associated with the publication of research findings in peer-reviewed scientific journals. Issues included open access vs. paywall publication, 'alternatives to publication in a traditional academic journal' (The Report, p. 35, paragraph 97), the mandating of greater data transparency and data deposition by scientific journals, the encouragement of data management plans by funders, the encouragement of the 'Publication of replications and null

results' (The Report, p. 38), and an effective process for the correction of scientific articles and the retraction of findings that are flawed or whose reliability is questionable.

The sixth chapter ('Solving reproducibility challenges in the assessment of research', The Report, p. 42–48, paragraphs 124–144) focused on a change in the academic reward system away from indicators such as number of publications, the prestige of journals where findings are published, and the summing of research income, towards open and transparent research practices and broader contributions to the academic field including 'voluntary peer review and promoting reproducibility and research integrity' (The Report, p. 45, paragraph 133). Moreover, the chapter argued for more rigorous peer review systems and a 'registered report' model, in which the study design is peer-reviewed and the subsequent articles are published independently of the nature of the findings (The Report, p. 47, paragraph 140). Chapter six concluded: '144. We hope that this Report has raised practicable and effective solutions to the challenges presented by a lack of reproducibility in research. As we have argued, maximising the integrity and utility of the research system is a vital element of realising the UK's science superpower ambitions and, if realised, will deliver great benefits and avert significant risks across the board'. (The Report, p. 48)

4. Suggested Issues and Solutions from The Report are Consistent with those Widely Found in the Scientific Literature

The reasons provided for the reproducibility crisis (methodological/statistical shortcomings, a focus on novelty resulting in publication bias/file drawer problem, a publish or perish culture, a lack of data transparency, insufficient scrutiny/thoroughness by reviewers) and the solutions proposed (better training, more control, improved guidelines for conduct and data provision, promoting replication studies and the publication of null results, an incentive system that rewards reproducible research) align with the discourse in the scientific literature (Ioannidis 2005; Alberts et al. 2015; Begley and Ioannidis 2015; Benjamin et al. 2018; Catillon 2019; Goldacre, Morton, and DeVito 2019; Laraway et al. 2019; Wass, Ray, and Michaelis 2019; Munafò et al. 2020; Munafò et al. 2020; Diaba-Nuhoho and Amponsah-Offeh 2021; Stewart et al. 2021; UK Reproducibility Network Steering Committee 2021; Freese, Rauf, and Voelkel 2022; Macleod 2022; Oransky 2022; Rajtmajer, Errington, and Hillary 2022; Stewart et al. 2022; Wang and Long 2022; Ball 2023; Butler et al. 2023; Oransky, Marcus, and Abritis 2023; Song et al. 2023; Yang et al. 2023; Oransky and Redman 2024).

Notably, key academic stakeholders and organisations (e.g. the UK Reproducibility Network, the UK Reproducibility Network Steering Committee, Retraction Watch) contributed both to the relevant scientific literature and in evidence to The Report (Munafò et al. 2017; Goldacre, Morton, and DeVito 2019; Munafò et al. 2020; Stewart et al. 2021; UK Reproducibility Network Steering Committee 2021; MacLeod et al. 2022; Oransky 2022; Stewart et al. 2022; Butler et al. 2023; Oransky, Marcus, and Abritis 2023; Oransky and Redman 2024, The Report).

5. Lack of Focus and of Accuracy and Precision in Terminology

Although the content of The Report largely reflects the scientific literature, it is affected by shortcomings in scope and in the accuracy and precision of its terminology. For

example, paragraph 98 formulates the aim that 100% of scientific outputs will be published as open access 'by the end of 2025 at the latest' (The Report, p.36). Although this aim may well be desirable to deliver the goal of equitable access to data and evidence, open access does not impact on data reproducibility and is hence not relevant in this context.

The Report's use of the term 'reproducibility' is very general and unspecific: 'Therefore, in line with the evidence we heard, we use the term reproducibility to broadly refer to the transparency and quality of research—i.e. have the researchers been sufficiently open to allow for robust assessment of their work's data, methodology, and conclusions, and, through such an assessment, is their research proven to be thorough and accurate'. (The Report, p. 7/8, paragraph 9)

The importance of a level of transparency that enables others to repeat research is without doubt essential, in particular for the analysis of large, complex datasets whose analysis can lead to differing and even conflicting results (Bogner et al. 2011; Botvinik-Nezer et al. 2020; Schweinsberg et al. 2021; Gould et al. 2023), including meta-analyses (Bar-Yam et al. 2023; Jefferson et al. 2023; Greenhalgh et al. 2024). There is also no reasonable argument that could be made against the full transparency of research (apart from data protection and privacy issues). However, using the term 'reproducibility' for the 'transparency and quality of research' across different disciplines (The Report, p. 7/8, paragraph 9) is misleading and inaccurate.

In the computer sciences, 'reproducibility' often refers to the documentation of a study that enables someone else to perform exactly the same analyses and to receive the same results (Piccolo and Frampton 2016). In such contexts (in computer sciences and beyond), where different researchers apply identical methods on identical datasets, nobody can disagree that exact repeats have to provide identical results. Considering the second part of the above definition of reproducibility ('have the researchers been sufficiently open to allow for robust assessment of their work's data, methodology, and conclusions, and, through such an assessment, is their research proven to be thorough and accurate', The Report, p. 7/8, paragraph 9), identical results from the analysis of the same datasets by the same method provide axiomatic confirmation that 'the researchers' have 'been sufficiently open' and 'thorough'. However, even in such a scenario, identical results do not show (still less 'prove') that the research is 'accurate'. Large-scale studies, in which many different researchers and research groups analysed the same dataset, resulted in broad ranges of differing results, including contradictory ones (Schweinsberg et al. 2021; Gould et al. 2023). In these studies, every one of the analyses can be presented in a '*reproducible*' manner that enables others to repeat exactly the same analysis and to obtain exactly the same results. However, not all of the differing reproducible results can be 'accurate' or 'true', thereby providing immediate evidence for the conceptual and practical disambiguation of 'reproducibility' from 'accuracy', 'truth', and 'proof'.

In the experimental life sciences, the situation is even more complex. Here, exact repeats can result in varying outcomes that can cover a remarkable range of values (Rudeck et al. 2020; Suman and Lino de Oliveira 2022; Reddin et al. 2023). For example, the NCI60 is a project of the National Cancer Institute in the US National Institutes of Health (NIH) that has repeatedly tested different compounds including approved anti-cancer drugs in the same cancer cell lines since the 1980s, applying the highest

available quality standards (Shoemaker 2006; Shankavaram et al. 2009; Reinhold et al. 2012; Luna et al. 2021). A recent analysis of the resulting dataset determined the variability among independent experiments testing the same compounds in the same cell lines (Reddin et al. 2023). The differences (fold changes) between the lowest and the highest GI50 (the concentration of a compound that reduces cell growth by 50%) for individual compound/cell line combinations increased with the number of experiments. The maximum fold change observed for a compound/cell line combination was 3.16×10^{10}. All experiments that had been performed more than 100 times had fold changes greater than five, and 70.5% of these experiments displayed fold changes greater than 1,000 (Reddin et al. 2023). These are large levels of variation, given that anti-cancer drugs are usually administered at a maximum dose that cannot be further increased in cancer patients due to toxicity and/or whose efficacy does not increase at a higher dose, because the applied dose causes maximum inhibition of the respective drug target (Eisenhauer et al. 2000; Sachs et al. 2016; Corbaux et al. 2019; Mansinho et al. 2019). Such a large level of variation among relatively simple cell culture experiments (Reddin et al. 2023) and high levels of variation among more complex animal experiments (Suman and Lino de Oliveira 2022) indicate the limits of standardisation in experimental research (Wass, Ray, and Michaelis 2019; Suman and Lino de Oliveira 2022).

Rather obviously, standardisation is often not an option in research, as research is often performed for the first time, using novel (model) systems and/or novel approaches in the absence of comparative data and without the possibility of standardisation (Wass, Ray, and Michaelis 2019; Penders, de Rijcke, and Holbrook 2020). Even if the standardisation of an approach is feasible, it is only robust if the 'true' or 'accurate' answer is known. Otherwise, standardisation can result in the consistent reproduction of a 'false' or 'inaccurate' finding for, as noted above, a technique or an experiment can be precise and reproducible without providing accurate results (Draghici et al. 2006).

The limitations of standardisation and of performing experiments that produce 'accurate' results are illustrated by the limitations of diagnostic tests. Despite substantial effort, the availability of reliable standards, and the regular performance of interlaboratory comparisons, sensitivity (true positive) and specificity (true negative) rates never reach 100%, i.e. there are always false positives and false negatives (Akobeng 2007). Hence, a single experiment and a single study can never provide a definitive answer to a research question. There is always a risk that even a study performed to the highest standards produces an outlier finding due to the variability in the measurement. The probabilistic nature of quantitative research findings can result in false positives or false negatives, given that the *p*-value is the result of a sampling statistic, which can be subject to sampling errors. For example, if 1,000 hypotheses are tested of which 100 are true, among the 900 false ones, 45 can be expected to be false positive when a 5% threshold is applied (Forstmeier, Wagenmakers, and Parker 2017). Therefore, every research study should be considered merely as a contribution to the ongoing scientific investigation. This has long been appreciated by the concept of self-correction, i.e. that findings that do not hold up to subsequent scrutiny are corrected or at least disappear into oblivion, because future research can only build on previous findings that can be reliably confirmed, be it by direct replication and/ or by complementary approaches testing the same hypothesis (Popper 1963; Merton 1973; Alger 2020; Bernard 2023).

Taken together, it is important to recognise and appreciate that even a study performed following the highest standards and applying the highest levels of integrity (i.e. research has been performed and reported in an 'open' and 'accurate' way) may not provide an 'accurate' or 'true' result. Hence, we consider that the definition and use of the term 'reproducibility' for research that is 'proven to be thorough and accurate' (The Report, p. 7/8, paragraph 9) sets unrealistic expectations by misrepresenting the way in which knowledge is generated, i.e. by multiple complementary studies over time. Notably, the lack of clarity and a clear definition of what is meant by 'reproducibility' in a certain context is a general issue in the literature on the topic and is not unique to The Report. Lack of precision and vague terminology have been identified as obstacles to the development of concrete, constructive approaches for the meaningful investigation of different forms of data reproducibility (Goodman, Fanelli, and Ioannidis 2016; Simkus et al. 2025), and a precise, unambiguous, and applicable terminology remains needed to identify concrete problems that can be investigated in a constructive and focused way.

6. Lack of Knowledge about the Existence and Extent of the 'Reproducibility Crisis' Prevents the Development of Measurable Success Criteria

The Report itself refers in different places to the lack of (in particular quantitative) evidence on whether a reproducibility crisis exists and, if it does, on the size of the issue. This starts with the following statement in the summary: 'We found that while there are many reports of problems of non reproducibility, there has been no comprehensive and rigorous assessment of the scale of the problem in the UK, nor which disciplines are most affected and therefore the extent to which this is indeed a 'crisis''. (The Report, p. 3)

Moreover, paragraph 27 deals with the lack of quantitative evidence and how this knowledge gap affects the design of meaningful measures to address a potential lack of data quality: 'Although qualitative evidence indicates a potentially substantial scale of research integrity issues in the UK, there is a lack of quantitative evidence, including on the relative significance of the different causes of problems. This can only hamper efforts to evaluate damage being caused to the UK research sector in terms of culture, performance, reputation and economic value—now and in the future. This in turn prevents the design of proportionate and effective solutions to any problems'. (The Report, p. 15)

Finally, the 'Conclusions and recommendations' section starts with the following statement: '1. Although qualitative evidence indicates a potentially substantial scale of research integrity issues in the UK, there is a lack of quantitative evidence, including on the relative significance of the different causes of problems. This can only hamper efforts to evaluate damage being caused to the UK research sector in terms of culture, performance, reputation and economic value—now and in the future. This in turn prevents the design of proportionate and effective solutions to any problems'. (The Report, p. 49)

This lack of evidence on the existence and size of the 'reproducibility crisis' is also appreciated in the literature. Despite the prominence of the topic and its fundamental importance, evidence is largely anecdotal and based on researcher beliefs, often expressed

in survey responses or published as 'Comments or Correspondence' without providing detailed information (Prinz, Schlange, and Asadullah 2011; Begley and Ellis 2012; Mobley et al. 2013; Baker 2016; Nature Editorial 2018; Hopp and Hoover 2019; Wass, Ray, and Michaelis 2019; Farrar, Ostojić, and Clayton 2021; Samuel and König-Ries 2021; de Ridder 2022; Hicks 2023; Calnan et al. 2024). Actual reproducibility studies are rarely performed (Wass, Ray, and Michaelis 2019; de Ridder 2022; Roper et al. 2022; Calnan et al. 2024), and if they are, their interpretation can be controversial (Wass, Ray, and Michaelis 2019; Mullard 2021; O'Grady 2023; Calnan et al. 2024). Notably, there are also voices that question the crisis narrative (Fanelli 2018; Imam 2022; Moody, Keister, and Ramos 2022, Penders, de Rijcke, and Holbrook 2020; Schauer 2023; Schneck 2023).

The lack of agreed criteria indicating a successful replication is illustrated by unresolved scientific disputes. For example, the 'Reproducibility Project: Cancer Biology' (Center for Open Science 2021) selected 50 experiments from 23 highly influential preclinical cancer studies published between 2010 and 2012 for independent replication by a project team. The authors concluded that only five of the investigated studies were successfully replicated (Errington et al. 2021). However, this is a limited dataset focused on small, early, highly cited studies, which are known to be more likely to overestimate effects (Fanelli, Costas, and Ioannidis 2017), and may not be representative of the reproducibility of experimental life science research in general. Notably, the conclusion of the authors is itself subject of dispute. Authors of reports that were considered not successfully replicated by the 'Reproducibility Project: Cancer Biology' claimed that their findings had been independently confirmed by other groups in the meantime and had resulted in clinical drug candidates currently undergoing clinical testing (Mullard 2021).

There is also an unresolved dispute about the consistency of the results of two large pharmacogenomic screens, the Genomics of Drug Sensitivity in Cancer database and the Cancer Cell Line Encyclopedia (Barretina et al. 2012; Garnett et al. 2012). Both studies had tested large panels of anti-cancer drugs in large numbers of cancer cell lines. When different research groups compared the data from both studies, they could not agree on whether the observed differences were within an acceptable, expected range or not (Haibe-Kains et al. 2013; Cancer Cell Line Encyclopedia Consortium & Genomics of Drug Sensitivity in Cancer Consortium 2015; Bouhaddou et al. 2016; Geeleher et al. 2016; Mpindi et al. 2016; Safikhani et al. 2016a, 2016b, 2016c, 2016d, 2016e). Such unresolved disputes are consistent with findings showing that reported failure rates can be subjective, biased, and arbitrary and that low or high failure rates can be the consequence of chance (Schauer 2023). Moreover, replication studies in the life sciences under highly standardised conditions (including the data from the NCI60 study outlined above in chapter 4) reported a high level of inherent experimental variability suggesting that accounts of limited reproducibility may be based (at least in part) on unrealistic expectations of the technically feasible level of data reproducibility (Karp et al. 2014; Wass, Ray, and Michaelis 2019; Rudeck et al. 2020; Suman and Lino de Oliveira 2022; Reddin et al. 2023; Calnan et al. 2024).

Considering the very influential Nature survey from 2016 (Baker 2016) in more detail, while 70% of the 1,576 respondents reported to have failed to reproduce someone else's results at least once, more than 50% of the respondents had also failed at least once to replicate their own findings (Baker 2016). It seems reasonable to assume that the

survey respondents did not purposely produce data that they themselves could not reproduce. Moreover, it is plausible that it is more difficult to reproduce somebody else's than your own data. Consequently, the relative difference in replication failures reported between own and other scientists' experiments does not appear too dramatic, which further supports the notion that the perception of a 'reproducibility crisis' may be (at least in part) the consequence of expectations that are based on intuition rather than on actual data. Taken together, The Report rightly highlights a lack of evidence on the existence and scale of the alleged 'reproducibility crisis'. However, instead of drawing the obvious conclusion that, in the absence of evidence, the next step has to be the establishment of reliable evidence, The Report proposes a wide range of measures (see next section) to tackle a problem whose existence remains questionable and unsupported.

7. Exact Nature and Outcomes of Proposed Measures Remain Vague, Elusive, and Unmeasurable

From page 49 onwards, The Report outlines 28 'Conclusions and recommendations'. Under the headline 'The extent and impact of reproducibility challenges within UK research', the first six conclusions and recommendations are of a political and organisational nature and do not directly concern research practices (The Report, p. 49). Recommendations 17–23 ('Solving reproducibility challenges which emerge in the communication of research', The Report, p. 51) focus on the transparent provision of data and all relevant information that enable the reproduction of experiments and studies. As mentioned before, maximum transparency is without doubt highly desirable.

The sections 'Solving reproducibility challenges which emerge as research is conducted' (The Report, p. 50/51, recommendations 7 to16) and 'Solving reproducibility challenges in the assessment of research' (The Report, p. 52, recommendations 24–28) contain recommendations on how to improve the research environment to increase researcher integrity and, in turn, data reproducibility. A number of the recommendations focus on the research culture, including training, funding, and the reward/incentive systems. This involves recommendations on how to deal with research misconduct (The Report, recommendations 7,8) and on a research environment that promotes 'research integrity and reproducibility' by (1) putting more emphasis on 'reproducibility' instead of 'originality' in the evaluation of researchers; (2) providing training in research integrity and reproducibility to undergraduate and postgraduate students as well as to academic staff; (3) providing more time for reproducible research; (4) reducing pressure and increasing security for researchers by providing at least three-year research contracts for postdoctoral researchers and protecting research time allocations for academic researchers; (5) reducing pressure to secure funding and produce frequent outputs in prestigious journals; and (6) considering broader researcher contributions 'including time spent conducting voluntary peer review and promoting reproducibility and research integrity' in assessments (The Report, recommendations 9,10,13,14,15,16,24,25,26).

Additional recommendations include more scrutiny during the peer review process, including the publication of reviewer reports and authors' responses (The Report, recommendation 27) and the use of registered reports, in which the decision about the publication of a study by a journal is based on the peer review of the methodology prior to its performance and independently of the actual results (The Report, recommendation 28).

Finally, the funding and provision of career paths for statisticians and software developers (The Report, recommendation 11) and the implementation of 'stronger tests for the presence of adequate software and statistical skills' among grant applicants and a 'methodological support system' (The Report, recommendation 12) are recommended.

In places, the recommendations are affected by a lack of focus that is the consequence of the conflation of 'reproducibility' and 'integrity', as outlined in chapter 4, above. For example, recommendation 17 formulates the aim that, 'UKRI and other research funders should continue to implement open access policies until this figure reaches 100%, by the end of 2025 at the latest'. Open access removes barriers to the availability of data and information, but it cannot in itself increase the success rates of replication experiments. Similarly, the timely removal/retraction of articles reporting research results that cannot be reproduced (The Report, recommendation 23), does not improve data quality per se.

Most importantly, there is currently no way to measure the success of the proposed changes. As laid out in the preceding section, 5, it is not clear whether a 'reproducibility crisis' exists that goes significantly beyond inherent experimental variation, and if it exists, its size cannot be determined (Karp et al. 2014; Fanelli 2018; Wass, Ray, and Michaelis 2019; Rudeck et al. 2020; Imam 2022; Penders, de Rijcke, and Holbrook 2020; Suman and Lino de Oliveira 2022; Reddin et al. 2023; Schauer 2023; Schneck 2023; Calnan et al. 2024). Consequently, we do not know whether there is a reproducibility crisis that needs addressing or, if there is one, how it is possible to determine and then to assess the success of measures proposed to improve data quality.

It would be more appropriate to address the alleged reproducibility crisis in an evidence-based approach that focuses on establishing whether there is a problem and, if there is, what its scale actually amounts to. Assessment of scale requires more diligent attention to the potential scope of the alleged crisis than has so far been provided, including careful consideration of experimental and non-experimental social sciences and of the high-value topics most often subject to reproducibility concerns in those areas — for example, in applied clinical psychology, mental disorder, learning, neuropsychology, and crime (Moody, Keister, and Ramos 2022, 68–69). Emerging research in sociological areas bearing directly on policy and legislation suggests that considerable specificity and nuance applies (e.g. in socio-legal scholarship: Cahill-O'Callaghan and Mulcahy 2022). An evidence-based approach across the natural and social sciences must focus on determining the inherent data variability of research data in each area. This will require the collection of data in multiple, complementary ways. The example of the analysis of NCI60 panel data (Reddin et al. 2023) illustrates how existing datasets can be used to develop an improved understanding of the variability of experimental data. Such an approach could be extended to other datasets. Moreover, data variability could be prospectively studied by repeatedly performing the same experiment, ideally with variations in parameters such as the reagents and experimentalists and over prolonged times. This could happen in (large) dedicated studies and could become a routine aspect of all research projects. In short, all raw data considered to fulfil common quality standards should be collected, made available, and analysed. Such a combined approach would over time result in a much more in-depth understanding of the inherent variability of research data across the disciplines.

Surprisingly, 'publication bias', the confounder of data quality that is supported by evidence and that could be addressed, is not explicitly mentioned in the recommendations

at the end of The Report, although it is extensively discussed in the section 'Publication of replications and null results' (The Report, p. 38, paragraphs 107–116). Publication bias refers to the phenomenon that significant ('positive') findings are more likely to be published than non-significant ('negative') ones ('file drawer problem') (Sterling 1959; Rosenthal 1979; Begley and Ioannidis 2015; Wass, Ray, and Michaelis 2019; Wieschowski et al. 2019; Marks-Anglin and Chen 2020; Herbet et al. 2022). As a consequence, the published literature is biased towards positive results and can be misleading, because negative findings go unpublished. This problem could be addressed in an evidence-based approach. There could be an expectation that data derived from grant-funded projects are at least disseminated in yet-to-design databases or as preprints, if not via peer-reviewed research articles. Adherence could be monitored as exemplified for clinical trials. Notably, adherence to the requirements for the reporting of clinical trial results is still limited (DeVito, Bacon, and Goldacre 2020; Nelson et al. 2023). Hence, additional incentives may be needed. For example, the availability of data from previous grant-funded projects could be (within means) a criterion for the assessment of future grant proposals to encourage adherence to such guidelines.

8. Conclusion

The 'Sixth Report—Reproducibility and Research Integrity' has substantial shortcomings in focus and logic. The very broad use of the term 'reproducibility' in a way that also includes 'research integrity' is flawed. These are separate issues, although there is an overlap between them. Not all questionable research practices affect reproducibility, e.g. 'honorary' or 'ghost' authorships (Justin et al. 2022; Pruschak and Hopp 2022).

However, the main shortcoming of the 'Sixth Report—Reproducibility and Research Integrity' is that it addresses a problem (the 'reproducibility crisis') that may not exist, an issue that is appreciated numerous times in The Report itself. If there is a 'reproducible crisis', its scale is unknown. Therefore, the next evidence-based step has to be the establishment of an understanding of the inherent variability of research data across disciplines and research settings. This requires that the provision of all high-quality research data becomes the norm and a routine part of all research. These data can then be systematically analysed, which will over time result in an improved understanding of the technically feasible level of data reproducibility and repeatability. Only an understanding of the inherent variability in research data will enable the identification of instances where the variability exceeds the technically feasible level of variation, and, where this is the case, facilitate the assessment of the impact of any remedial actions then taken. Without such knowledge the implementation of measures remains unsupported and premature, with The Report's lack of an evidence-based approach towards the research environment diverting from the scrutiny that is normally expected and applied to the scientific method. Hence, our findings are a reminder that the research environment and culture require the same level of attention and rigour as the experimental design itself.

Acknowledgements

Not applicable.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Martin Michaelis http://orcid.org/0000-0002-5710-5888

References

Akobeng, A. K. 2007. "Understanding Diagnostic Tests 1: Sensitivity, Specificity and Predictive Values." *Acta Paediatrica* 96 (3): 338–341. https://doi.org/10.1111/j.1651-2227.2006.00180.x.

Alberts, B., R. J. Cicerone, S. E. Fienberg, A. Kamb, M. McNutt, R. M. Nerem, R. Schekman, et al. 2015. "Scientific Integrity. Self-correction in Science at Work." *Science* 348 (6242): 1420–1422. https://doi.org/10.1126/science.aab3847.

Alger, B. E. 2020. "Scientific Hypothesis-Testing Strengthens Neuroscience Research." *eNeuro* 7 (4): ENEURO.0357-19.2020. https://doi.org/10.1523/ENEURO.0357-19.2020.

Baker, M. 2016. "1,500 Scientists Lift the lid on Reproducibility." *Nature* 533 (7604): 452–454. https://doi.org/10.1038/533452a.

Ball, P. 2023. "Is AI Leading to a Reproducibility Crisis in Science?" *Nature* 624 (7990): 22–25. https://doi.org/10.1038/d41586-023-03817-6.

Bar-Yam, Y., J. M. Samet, A. F. Siegenfeld, and N. N. Taleb. 2023. Quantitative errors in the Cochrane review on "Physical interventions to interrupt or reduce the spread of respiratory viruses". arXiv [Preprint]. 2023:2310.15198. https://doi.org/10.48550/arXiv.2310.15198.

Barretina, J., G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, et al. 2012. "The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity." *Nature* 483 (7391): 603–607. https://doi.org/10.1038/nature11003.

Begley, C. G., and L. M. Ellis. 2012. "Drug Development: Raise Standards for Preclinical Cancer Research." *Nature* 483 (7391): 531–533. https://doi.org/10.1038/483531a.

Begley, C. G., and J. P. Ioannidis. 2015. "Reproducibility in Science: Improving the Standard for Basic and Preclinical Research." *Circulation Research* 116 (1): 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819.

Benjamin, D. J., J. O. Berger, M. Johannesson, B. A. Nosek, E. J. Wagenmakers, R. Berk, K. A. Bollen, et al. 2018. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1): 6–10. https://doi.org/10.1038/s41562-017-0189-z.

Bernard, C. 2023. "Stop Reproducing the Reproducibility Crisis." *eNeuro* 10 (2): ENEURO.0032-23.2023. https://doi.org/10.1523/ENEURO.0032-23.2023.

Bogner, V., B. A. Leidel, K. G. Kanz, W. Mutschler, E. A. Neugebauer, and P. Biberthaler. 2011. "Pathway Analysis in Microarray Data: A Comparison of two Different Pathway Analysis Devices in the Same Data set." *Shock* 35 (3): 245–251. https://doi.org/10.1097/SHK.0b013e3181fc904d.

Botvinik-Nezer, R., F. Holzmeister, C. F. Camerer, A. Dreber, J. Huber, M. Johannesson, M. Kirchler, et al. 2020. "Variability in the Analysis of a Single Neuroimaging Dataset by Many Teams." *Nature* 582 (7810): 84–88. https://doi.org/10.1038/s41586-020-2314-9.

Bouhaddou, M., M. S. DiStefano, E. A. Riesel, E. Carrasco, H. Y. Holzapfel, D. C. Jones, G. R. Smith, et al. 2016. "Drug Response Consistency in CCLE and CGP." *Nature* 540 (7631): E9–E10. https://doi.org/10.1038/nature20580.

Butler, J., N. Jacobs, M. Macleod, and M. Munafò. 2023. "The UK Reproducibility Network: A Progress Report." *Journal of Neuroscience Methods* 397:109949. https://doi.org/10.1016/j.jneumeth.2023.109949.

Cahill-O'Callaghan, R., and L. Mulcahy. 2022. "Where Are the Numbers? Challenging the Barriers to Quantitative Socio-Legal Scholarship in the United Kingdom." *Journal of Law and Society* 49: s105–s118. https://doi.org/10.1111/jols.12376.

Calnan, M., S. Kirchin, D. L. Roberts, M. N. Wass, and M. Michaelis. 2024. "Understanding and Tackling the Reproducibility Crisis – Why We Need to Study Scientists' Trust in Data." *Pharmacological Research* 199:107043. https://doi.org/10.1016/j.phrs.2023.107043.

Cancer Cell Line Encyclopedia Consortium; Genomics of Drug Sensitivity in Cancer Consortium. 2015. "Pharmacogenomic Agreement between two Cancer Cell Line Data Sets." *Nature* 528 (7580): 84–87. https://doi.org/10.1038/nature15736.

Catillon, M. 2019. "Trends and Predictors of Biomedical Research Quality, 1990-2015: A Meta-research Study." *BMJ Open* 9 (9): e030342. https://doi.org/10.1136/bmjopen-2019-030342.

Center for Open Science. 2021. "Reproducibility Project Cancer Biology." https://www.cos.io/rpcb on 3 May 2024.

Corbaux, P., M. El-Madani, M. Tod, J. Péron, D. Maillet, J. Lopez, G. Freyer, and B. You. 2019. "Clinical Efficacy of the Optimal Biological Dose in Early-Phase Trials of Anti-cancer Targeted Therapies." *European Journal of Cancer* 120:40–46. https://doi.org/10.1016/j.ejca.2019.08.002.

de Ridder, J. 2022. "How to Trust a Scientist." *Studies in History and Philosophy of Science* 93:11–20. https://doi.org/10.1016/j.shpsa.2022.02.003.

DeVito, N. J., S. Bacon, and B. Goldacre. 2020. "Compliance with Legal Requirement to Report Clinical Trial Results on ClinicalTrials.gov: A Cohort Study." *The Lancet* 395 (10221): 361–369. https://doi.org/10.1016/S0140-6736(19)33220-9.

Diaba-Nuhoho, P., and M. Amponsah-Offeh. 2021. "Reproducibility and Research Integrity: The Role of Scientists and Institutions." *BMC Research Notes* 14 (1): 451. https://doi.org/10.1186/s13104-021-05875-3.

Draghici, S., P. Khatri, A. C. Eklund, and Z. Szallasi. 2006. "Reliability and Reproducibility Issues in DNA Microarray Measurements." *Trends in Genetics* 22 (2): 101–109. https://doi.org/10.1016/j.tig.2005.12.005.

Eisenhauer, E. A., P. J. O'Dwyer, M. Christian, and J. S. Humphrey. 2000. "Phase I Clinical Trial Design in Cancer Drug Development." *Journal of Clinical Oncology* 18 (3): 684–692. https://doi.org/10.1200/JCO.2000.18.3.684.

Errington, T. M., M. Mathur, C. K. Soderberg, A. Denis, N. Perfito, E. Iorns, and B. A. Nosek. 2021. "Investigating the Replicability of Preclinical Cancer Biology." *Elife* 10:e71601. https://doi.org/10.7554/eLife.71601.

Fanelli, D. 2018. "Opinion: Is Science Really Facing a Reproducibility Crisis, and Do We Need It to?" *Proceedings of the National Academy of Sciences* 115 (11): 2628–2631. https://doi.org/10.1073/pnas.1708272114.

Fanelli, D., R. Costas, and J. P. Ioannidis. 2017. "Meta-assessment of Bias in Science." *Proceedings of the National Academy of Sciences* 114 (14): 3714–3719. https://doi.org/10.1073/pnas.1618569114.

Farrar, B. G., L. Ostojić, and N. S. Clayton. 2021. "The Hidden Side of Animal Cognition Research: Scientists' Attitudes toward Bias, Replicability and Scientific Practice." *PLoS One* 16 (8): e0256607. https://doi.org/10.1371/journal.pone.0256607.

Forstmeier, W., E. J. Wagenmakers, and T. H. Parker. 2017. "Detecting and Avoiding Likely False-Positive Findings – a Practical Guide." *Biological Reviews* 92 (4): 1941–1968. https://doi.org/10.1111/brv.12315.

Freese, J., T. Rauf, and J. G. Voelkel. 2022. "Advances in Transparency and Reproducibility in the Social Sciences." *Social Science Research* 107:102770. https://doi.org/10.1016/j.ssresearch.2022.102770.

Garnett, M. J., E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P. Greninger, et al. 2012. "Systematic Identification of Genomic Markers of Drug Sensitivity in Cancer Cells." *Nature* 483 (7391): 570–575. https://doi.org/10.1038/nature11005.

Geeleher, P., E. R. Gamazon, C. Seoighe, N. J. Cox, and R. S. Huang. 2016. "Consistency in Large Pharmacogenomic Studies." *Nature* 540 (7631): E1–E2. https://doi.org/10.1038/nature19838.

Goldacre, B., C. E. Morton, and N. J. DeVito. 2019. "Why Researchers Should Share Their Analytic Code." *BMJ* 367:l6365. https://doi.org/10.1136/bmj.l6365.

Goodman, S. N., D. Fanelli, and J. P. Ioannidis. 2016. "What Does Research Reproducibility Mean?" *Science Translational Medicine* 8 (341): 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027.

Gould, E., H. S. Fraser, T. H. Parker, S. Nakagawa, S. C. Griffith, P. A. Vesk, F. Fidler, et al. 2023. "Same Data, Different Analysts: Variation in Effect Sizes due to Analytical Decisions in Ecology and Evolutionary Biology." *EcoevoRxiv*, https://doi.org/10.32942/X2GG62.

Greenhalgh, T., C. R. MacIntyre, M. G. Baker, S. Bhattacharjee, A. A. Chughtai, D. Fisman, M. Kunasekaran, A. Kvalsvig, D. Lupton, M. Oliver, et al. 2024. "Masks and Respirators for Prevention of Respiratory Infections: A State of the Science Review." *Clinical Microbiology Reviews* 37 (2): e0012423. https://doi.org/10.1128/cmr.00124-23.

Haibe-Kains, B., N. El-Hachem, N. J. Birkbak, A. C. Jin, A. H. Beck, H. J. Aerts, and J. Quackenbush. 2013. "Inconsistency in Large Pharmacogenomic Studies." *Nature* 504 (7480): 389–393. https://doi.org/10.1038/nature12831.

Herbet, M. E., J. Leonard, M. G. Santangelo, and L. Albaret. 2022. "Dissimulate or Disseminate? A Survey on the Fate of Negative Results." *Learned Publishing* 35 (1): 16–29. https://doi.org/10.1002/leap.1438.

Hicks, D. J. 2023. "Open Science, the Replication Crisis, and Environmental Public Health." *Accountability in Research* 30 (1): 34–62. https://doi.org/10.1080/08989621.2021.1962713.

Hopp, C., and G. A. Hoover. 2019. "What Crisis? Management Researchers' Experiences with and Views of Scholarly Misconduct." *Science and Engineering Ethics* 25 (5): 1549–1588. https://doi.org/10.1007/s11948-018-0079-4.

House of Commons Science, Innovation and Technology Committee. 2021. "Reproducibility of Research Inquiry Launched." Accessed March 25, 2024. https://committees.parliament.uk/work/1433/reproducibility-and-research-integrity/news/156859/reproducibility-of-research-inquiry-launched/.

House of Commons Science, Innovation and Technology Committee. 2022. "Written Evidence." Accessed March 25, 2024. https://committees.parliament.uk/work/1433/reproducibility-and-research-integrity/publications/written-evidence/.

House of Commons Science, Innovation and Technology Committee. 2023. "Sixth Report – Reproducibility and Research Integrity." Accessed July 28, 2023. https://committees.parliament.uk/work/1433/reproducibility-and-research-integrity/publications/ ('The Report').

Imam, A. A. 2022. "Remarkably Reproducible Psychological (Memory) Phenomena in the Classroom: Some Evidence for Generality from Small-N Research." *BMC Psychology* 10 (1): 274. https://doi.org/10.1186/s40359-022-00982-7.

Ioannidis, J. P. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

Jefferson, T., L. Dooley, E. Ferroni, L. A. Al-Ansary, M. L. van Driel, G. A. Bawazeer, M. A. Jones, et al. 2023. "Physical Interventions to Interrupt or Reduce the Spread of Respiratory Viruses." *Cochrane Database of Systematic Reviews* 2023 (1): CD006207. https://doi.org/10.1002/14651858.CD006207.pub6.

Justin, G. A., S. C. Miller, B. Tsou, X. Li, B. Purt, M. J. Fliotsos, J. Zhao, S. E. Gardner, et al. 2022. "Ghost and Honorary Authorship in Ophthalmology: A Cross-Sectional Survey." *American Journal of Ophthalmology* 240:67–78. https://doi.org/10.1016/j.ajo.2022.02.012.

Karp, N. A., A. O. Speak, J. K. White, D. J. Adams, Hrabé de Angelis M, Y. Hérault, and R. F. Mott. 2014. "Impact of Temporal Variation on Design and Analysis of Mouse Knockout Phenotyping Studies." *PLoS One* 9 (10): e111239. https://doi.org/10.1371/journal.pone.0111239.

Laraway, S., S. Snycerski, S. Pradhan, and B. E. Huitema. 2019. "An Overview of Scientific Reproducibility: Consideration of Relevant Issues for Behavior Science/Analysis." *Perspectives on Behavior Science* 42 (1): 33–57. https://doi.org/10.1007/s40614-019-00193-3.

Luna, A., F. Elloumi, S. Varma, Y. Wang, V. N. Rajapakse, M. I. Aladjem, J. Robert, C. Sander, Y. Pommier, and W. C. Reinhold. 2021. "CellMiner Cross-Database (CellMinerCDB) Version 1.2: Exploration of Patient-Derived Cancer Cell Line Pharmacogenomics." *Nucleic Acids Research* 49 (D1): D1083–D1093. https://doi.org/10.1093/nar/gkaa968.

Macleod, M., and University of Edinburgh Research Strategy Group. 2022. "Improving the Reproducibility and Integrity of Research: What Can Different Stakeholders Contribute?" *BMC Research Notes* 15 (1): 146. https://doi.org/10.1186/s13104-022-06030-2.

Mansinho, A., V. Boni, M. Miguel, and E. Calvo. 2019. "New Designs in Early Clinical Drug Development." *Annals of Oncology* 30 (9): 1460–1465. https://doi.org/10.1093/annonc/mdz191.
Marks-Anglin, A., and Y. Chen. 2020. "A Historical Review of Publication Bias." *Research Synthesis Methods* 11 (6): 725–742. https://doi.org/10.1002/jrsm.1452.
Merton, R. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*, 276. Chicago: University of Chicago Press.
Mobley, A., S. K. Linder, R. Braeuer, L. M. Ellis, and L. Zwelling. 2013. "A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic." *PLoS One* 8 (5): e63221. https://doi.org/10.1371/journal.pone.0063221.
Moody, J. W., L. A. Keister, and M. C. Ramos. 2022. "Reproducibility in the Social Sciences." *Annual Review of Sociology* 48 (1): 65–85. https://doi.org/10.1146/annurev-soc-090221-035954
Mpindi, J. P., B. Yadav, P. Östling, P. Gautam, D. Malani, A. Murumägi, A. Hirasawa, et al. 2016. "Consistency in Drug Response Profiling." *Nature* 540 (7631): E5–E6. https://doi.org/10.1038/nature20171
Mullard, A. 2021. "Half of top Cancer Studies Fail High-Profile Reproducibility Effort." *Nature* 600 (7889): 368–369. https://doi.org/10.1038/d41586-021-03691-0.
Munafò, M. R., C. D. Chambers, A. M. Collins, L. Fortunato, and M. R. Macleod. 2020. "Research Culture and Reproducibility." *Trends in Cognitive Sciences* 24 (2): 91–93. https://doi.org/10.1016/j.tics.2019.12.002.
Munafò, M. R., B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. P. du Sert, U. Simonsohn, E. J. Wagenmakers, J. J. Ware, and J. P. A. Ioannidis. 2017. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1 (1): 0021. https://doi.org/10.1038/s41562-016-0021.
Nature Editorial. 2018. "Checklists Work to Improve Science." *Nature* 556 (7701): 273–274. https://doi.org/10.1038/d41586-018-04590-7.
Nelson, J. T., T. Tse, Y. Puplampu-Dove, E. Golfinopoulos, and D. A. Zarin. 2023. "Comparison of Availability of Trial Results in ClinicalTrials.gov and PubMed by Data Source and Funder Type." *JAMA* 329 (16): 1404–1406. https://doi.org/10.1001/jama.2023.2351.
O'Grady, C. 2023. "Preregistering, Transparency, and Large Samples Boost Psychology Studies' Replication Rate to Nearly 90%." *Science*, https://doi.org/10.1126/science.adm8658.
Oransky, I. 2022. "Retractions Are Increasing, but Not Enough." *Nature* 608 (7921): 9. https://doi.org/10.1038/d41586-022-02071-6.
Oransky, I., A. Marcus, and A. Abritis. 2023. "How Bibliometrics and School Rankings Reward Unreliable Science." *BMJ* 382:p1887. https://doi.org/10.1136/bmj.p1887.
Oransky, I., and B. Redman. 2024. "Rooting out Scientific Misconduct." *Science* 383 (6679): 131. https://doi.org/10.1126/science.adn9352.
Penders, B., S. de Rijcke, and J. B. Holbrook. 2020. "Science's Moral Economy of Repair: Replication and the Circulation of Reference." *Accountability in Research* 27 (2): 107–113. https://doi.org/10.1080/08989621.2020.1720659.
Piccolo, S. R., and M. B. Frampton. 2016. "Tools and Techniques for Computational Reproducibility." *Gigascience* 5 (1): 30. https://doi.org/10.1186/s13742-016-0135-4.
Popper, K. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge. p. 293.
Prinz, F., T. Schlange, and K. Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" *Nature Reviews Drug Discovery* 10 (9): 712. https://doi.org/10.1038/nrd3439-c1.
Pruschak, G., and C. Hopp. 2022. "And the Credit Goes to … - Ghost and Honorary Authorship among Social Scientists." *PLoS One* 17 (5): e0267312. https://doi.org/10.1371/journal.pone.0267312.
Rajtmajer, S. M., T. M. Errington, and F. G. Hillary. 2022. "How Failure to Falsify in High-Volume Science Contributes to the Replication Crisis." *Elife* 11:e78830. https://doi.org/10.7554/eLife.78830.
Reddin, I. G., T. R. Fenton, M. N. Wass, and M. Michaelis. 2023. "Large Inherent Variability in Data Derived from Highly Standardised Cell Culture Experiments." *Pharmacological Research* 188:106671. https://doi.org/10.1016/j.phrs.2023.106671.

Reinhold, W. C., M. Sunshine, H. Liu, S. Varma, K. W. Kohn, J. Morris, J. Doroshow, and Y. Pommier. 2012. "CellMiner: A web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set." *Cancer Research* 72 (14): 3499–3511. https://doi.org/10.1158/0008-5472.CAN-12-1370.

Roper, K., A. Abdel-Rehim, S. Hubbard, M. Carpenter, A. Rzhetsky, L. Soldatova, and R. D. King. 2022. "Testing the Reproducibility and Robustness of the Cancer Biology Literature by Robot." *Journal of the Royal Society Interface* 19 (189): 20210821. https://doi.org/10.1098/rsif.2021.0821.

Rosenthal, R. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638–641. https://doi.org/10.1037/0033-2909.86.3.638.

Rudeck, J., S. Vogl, S. Banneke, G. Schönfelder, and L. Lewejohann. 2020. "Repeatability Analysis Improves the Reliability of Behavioral Data." *PLoS One* 15 (4): e0230900. https://doi.org/10.1371/journal.pone.0230900.

Sachs, J. R., K. Mayawala, S. Gadamsetty, S. P. Kang, and D. P. de Alwis. 2016. "Optimal Dosing for Targeted Therapies in Oncology: Drug Development Cases Leading by Example." *Clinical Cancer Research* 22 (6): 1318–1324. https://doi.org/10.1158/1078-0432.CCR-15-1295.

Safikhani, Z., N. El-Hachem, R. Quevedo, P. Smirnov, A. Goldenberg, N. Juul Birkbak, C. Mason, et al. 2016d. "Assessment of Pharmacogenomic Agreement." *F1000Research* 5:825. https://doi.org/10.12688/f1000research.8705.1.

Safikhani, Z., N. El-Hachem, P. Smirnov, M. Freeman, A. Goldenberg, N. J. Birkbak, A. H. Beck, et al. 2016a. "Reply." *Nature* 540 (7631): E2–E4. https://doi.org/10.1038/nature19839.

Safikhani, Z., N. El-Hachem, P. Smirnov, M. Freeman, A. Goldenberg, N. J. Birkbak, A. H. Beck, et al. 2016b. "Reply." *Nature* 540 (7631): E6–E8. https://doi.org/10.1038/nature20172.

Safikhani, Z., N. El-Hachem, P. Smirnov, M. Freeman, A. Goldenberg, N. J. Birkbak, A. H. Beck, et al. 2016c. "Reply." *Nature* 540 (7631): E11–E12. https://doi.org/10.1038/nature20581.

Safikhani, Z., P. Smirnov, M. Freeman, N. El-Hachem, A. She, Q. Rene, A. Goldenberg, et al. 2016e. "Revisiting Inconsistency in Large Pharmacogenomic Studies." *F1000Research* 5:2333. https://doi.org/10.12688/f1000research.9611.1.

Samuel, S., and B. König-Ries. 2021. "Understanding Experiments and Research Practices for Reproducibility: An Exploratory Study." *PeerJ* 9:e11140. https://doi.org/10.7717/peerj.11140.

Schauer, J. M. 2023. "On the Accuracy of Replication Failure Rates." *Multivariate Behavioral Research* 58 (3): 598–615. https://doi.org/10.1080/00273171.2022.2066500.

Schneck, A. 2023. "Are Most Published Research Findings False? Trends in Statistical Power, Publication Selection Bias, and the False Discovery Rate in Psychology (1975-2017)." *PLoS One* 18 (10): e0292717. https://doi.org/10.1371/journal.pone.0292717.

Schweinsberg, M., M. Feldman, N. Staub, O. R. van den Akker, R. C. M. van Aert, M. A. L. M. van Assen, Y. Liu, T. Althoff, et al. 2021. "Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis." *Organizational Behavior and Human Decision Processes* 165:228–249. https://doi.org/10.1016/j.obhdp.2021.02.003.

Shankavaram, U. T., S. Varma, D. Kane, M. Sunshine, K. K. Chary, W. C. Reinhold, Y. Pommier, and J. N. Weinstein. 2009. "CellMiner: A Relational Database and Query Tool for the NCI-60 Cancer Cell Lines." *BMC Genomics* 10 (1): 277. https://doi.org/10.1186/1471-2164-10-277.

Shoemaker, R. H. 2006. "The NCI60 Human Tumour Cell Line Anticancer Drug Screen." *Nature Reviews Cancer* 6 (10): 813–823. https://doi.org/10.1038/nrc1951.

Simkus, A., T. Coolen-Maturi, F. P. A. Coolen, and C. Bendtsen. 2025. "Statistical Perspectives on Reproducibility: Definitions and Challenges." *Journal of Statistical Theory and Practice* 19 (3): 40. https://doi.org/10.1007/s42519-025-00459-x.

Song, J., M. Solmi, A. F. Carvalho, J. I. Shin, and J. P. Ioannidis. 2023. "Twelve Years after the ARRIVE Guidelines: Animal Research Has Not Yet Arrived at High Standards." *Laboratory Animals* 58:109–115. https://doi.org/10.1177/00236772231181658.

Sterling, T. D. 1959. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance – or Vice Versa." *Journal of the American Statistical Association* 54 (285): 30–34. https://doi.org/10.2307/2282137.

Stewart, A. J., E. K. Farran, J. A. Grange, M. Macleod, M. Munafò, P. Newton, and D. R. Shanks. 2021. "Improving Research Quality: The View from the UK Reproducibility Network Institutional Leads for Research Improvement." *BMC Research Notes* 14 (1): 458. https://doi.org/10.1186/s13104-021-05883-3.

Stewart, S. L. K., C. R. Pennington, G. R. da Silva, N. Ballou, J. Butler, Z. Dienes, C. Jay, S. Rossit, and A. Samara. 2022. "Reforms to Improve Reproducibility and Quality Must Be Coordinated across the Research Ecosystem: The View from the UKRN Local Network Leads." *BMC Research Notes* 15 (1): 58. https://doi.org/10.1186/s13104-022-05949-w.

Suman, P. R., and C. Lino de Oliveira. 2022. "Systematic Heterogenisation to Improve Reproducibility in Animal Studies." *PLOS Biology* 20 (5): e3001629. https://doi.org/10.1371/journal.pbio.3001629.

UK Department for Business, Energy, and Industrial Strategy. 2022. "International Comparison of the UK Research Base, 2022." Accessed March 26, 2024. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1078073/international-comparison-uk-research-base-2022-accompanying-note.pdf.

UK Reproducibility Network Steering Committee. 2021. "From Grassroots to Global: A Blueprint for Building a Reproducibility Network." *PLoS Biology* 19 (11): e3001461. https://doi.org/10.1371/journal.pbio.3001461.

Wang, M., and Q. Long. 2022. "Addressing Common Misuses and Pitfalls of P Values in Biomedical Research." *Cancer Research* 82 (15): 2674–2677. https://doi.org/10.1158/0008-5472.CAN-21-2978.

Wass, M. N., L. Ray, and M. Michaelis. 2019. "Understanding of Researcher Behavior Is Required to Improve Data Reliability." *Gigascience* 8 (5): giz017. https://doi.org/10.1093/gigascience/giz017.

Wellcome Trust. 2020. "What Researchers Think about the Culture They Work In." Accessed March 26, 2024. https://wellcome.org/reports/What-researchers-think-about-research-culture.

Wieschowski, S., S. Biernot, S. Deutsch, S. Glage, A. Bleich, R. Tolba, and D. Strech. 2019. "Publication Rates in Animal Research. Extent and Characteristics of Published and non-published Animal Studies Followed up at Two German University Medical Centres." *PLoS One* 14 (11): e0223758. https://doi.org/10.1371/journal.pone.0223758.

Yang, Y., A. Sánchez-Tójar, R. E. O'Dea, D. W. A. Noble, J. Koricheva, M. D. Jennions, T. H. Parker, M. Lagisz, and S. Nakagawa. 2023. "Publication Bias Impacts on Effect Size, Statistical Power, and Magnitude (Type M) and Sign (Type S) Errors in Ecology and Evolutionary Biology." *BMC Biology* 21 (1): 71. https://doi.org/10.1186/s12915-022-01485-y.