



Understanding and tackling the reproducibility crisis – Why we need to study scientists' trust in data

Michael Calnan^a, Simon Kirchin^b, David L. Roberts^c, Mark N. Wass^d, Martin Michaelis^{d,e,*}

^a School of Social Policy, Sociology and Social Research, University of Kent, Canterbury, UK

^b Department of Philosophy, University of Kent, Canterbury, UK

^c Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, UK

^d School of Biosciences, University of Kent, Canterbury, UK

^e Dr Petra Joh Research Institute, Frankfurt am Main, Germany

ABSTRACT

In the life sciences, there is an ongoing discussion about a perceived 'reproducibility crisis'. However, it remains unclear to which extent the perceived lack of reproducibility is the consequence of issues that can be tackled and to which extent it may be the consequence of unrealistic expectations of the technical level of reproducibility. Large-scale, multi-institutional experimental replication studies are very cost- and time-intensive. This Perspective suggests an alternative, complementary approach: meta-research using sociological and philosophical methodologies to examine researcher trust in data. An improved understanding of the criteria used by researchers to judge data reliability will provide crucial, initial evidence on the actual scale of the reproducibility crisis and on measures to tackle it.

A "reproducibility crisis" (or "replication crisis") narrative is currently impacting the experimental life sciences (and other disciplines). Up to 90% of research funding has been suggested to be wasted due to a lack of reproducibility [1–3]. Despite the prominence of the topic, evidence is largely anecdotal and based on researcher beliefs, often expressed in survey responses or published as Comments or Correspondence without providing detailed information [1,2]. Actual reproducibility studies are rarely performed in the experimental life sciences [1,2], and if they are, their interpretation can be controversial [1,4].

The 'Reproducibility Project: Cancer Biology' was a collaboration between the Center for Open Science (<https://www.cos.io>) and Science Exchange (<https://www.scienceexchange.com/s/>) funded by the Laura and John Arnold Foundation (<https://www.cos.io/rpcb-contributors>). In this project, 50 experiments from 23 highly influential preclinical cancer studies published between 2010 and 2012 were selected for independent replication by a project team that also coordinated the performance of the replication experiments. According to the assessments of the authors, five of the investigated studies were successfully replicated [1,5]. However, all replication studies were performed by the same group of researchers. Moreover, this is a limited dataset focused on small, early, highly cited studies, which are known to be more likely to overestimate effects [1,6], and may not be representative of the reproducibility of experimental life science research in general. Notably, some 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 [4]. Hence, an agreed technically feasible level of data reproducibility does not exist [1,3–5].

Moreover, pilot studies indicated a substantial level of inherent experimental variability, including animal experiments performed under highly standardised conditions [1,7,8]. Our recent analysis of data from the US National Cancer Institute's NCI60 drug screening project indicated that the variation among experimental results is very high even under highly standardised conditions in cancer cell lines, which are a very simple biological model compared to animals. 70.5% of experiments that had been repeated at least 100 times displayed a > 1,000-fold difference between the lowest and the highest value [9].

Thus, it is not clear to which extent there is a reproducibility crisis in the experimental life sciences due to issues that can be addressed and to which extent the perception of a crisis is due to unrealistic expectations of the technical level of reproducibility.

However, there are indications that researchers develop a feeling for the trustworthiness of data in their respective fields. This was, for example, shown by the use of prediction markets [10,11]. In prediction markets, bets are placed on the outcomes of events and traded. When active psychology, economy, and social sciences researchers participated in a prediction market for the outcomes of replication studies they

* Corresponding author at: School of Biosciences, University of Kent, Canterbury, UK.

E-mail address: M.Michaelis@kent.ac.uk (M. Michaelis).

<https://doi.org/10.1016/j.phrs.2023.107043>

Received 9 October 2023; Received in revised form 12 December 2023; Accepted 18 December 2023

Available online 20 December 2023

1043-6618/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

forecasted the actual results with relatively high accuracy [10,11]. More than 90% of studies predicted not to replicate did not replicate, while 66% of studies predicted to replicate actually replicated [11]. The ability of researchers to judge (to some extent) data quality strongly suggests that there are factors in addition to the technical variability that affect reproducibility and that can be addressed to produce more reliable data. However, the scale of the issue and the nature of effective measures remain elusive.

The investigation of the (technically achievable) reproducibility levels across a comprehensive range of experimental systems in the life sciences would require an unprecedented investment of time and resources. Hence, we here suggest an alternative, complementary, interdisciplinary meta-research approach using sociological and philosophical methodologies to study researcher trust in data. This approach would provide initial, concrete evidence beyond anecdote about the suspected reproducibility crisis in a much shorter timeframe and at a fraction of the cost of extensive multi-institutional replication studies.

Trust between researchers is essential in the life sciences [2,12]. Researchers depend on findings from other researchers to develop their hypotheses and research questions. Moreover, most life sciences research is collaborative in nature. The resulting interdependencies cannot be maintained without a sufficient level of trust [2]. Additionally, trust enables researchers to overcome uncertainties about evidence. To be able to rely on someone's data, the trustor must have confidence in the competence, knowledge, and skills of the trustee, as well as confidence in their beneficence, fairness, and integrity [12,13]. Indeed, confidence in others' work enables researchers to trust their own work more as researchers are often building on others' results and methodologies.

Given the perception of a reproducibility crisis and a general lack of data reliability [1–3,6], asking on what basis researchers can still trust each other enough to continue their research in a meaningful way is a justified question. In this context, it has previously been suggested that scientists trust each other “in accordance with prevailing methodological standards” [2], which appreciates technical limitations and the real-world research culture that is not always ideal and can include questionable research practices like shortcuts and corner-cutting.

Hence, there is an underlying cultural knowledge that enables researchers to read between the lines in order to classify and judge the reliability of data beyond their literal presentation and to adjust for potential confounders, colloquially speaking to know when data need to be taken with a ‘pinch of salt’. This ability of researchers to judge data reliability is further supported by the previously mentioned results from prediction markets in other disciplines [10,11] and by findings showing that the focus of researchers when they read scientific articles shifts during career progression from the authors' data interpretation to the actual data, i.e. from introduction and discussion to the methods and results sections [14].

Consequently, investigating the (conscious and subconscious, explicit and tacit) criteria that enable researchers to interpret data, so that they can trust, interpret, and use them in a meaningful way, is an effective alternative strategy to learn more about the nature and extent of the suspected reproducibility crisis. This approach is complementary to and much less time-consuming than large scale experimental replication studies across a wide range of research fields whose feasibility will be limited by prohibitive costs [15]. If we learn how scientists handle and present their own data and how they can trust the research results of others, this will result in an improved understanding of the factors that shape the assessment of data quality.

Researcher trust in data includes trust in the process of data collection and analysis, in particular the transparency and completeness of the presented data (‘Open science’ [16]). Since data transparency is an issue that can and should be addressed under any circumstances as a matter of good practice, our proposed approach focuses on other aspects that may affect trust in data quality.

Notably, there are a number of approaches that could be used to establish an understanding of the (explicit and tacit) criteria that scientists use to assess data quality. We would envisage a mixed methods approach [17]. For example, a questionnaire of a random sample of life scientists enables the collection of representative population-based data. A questionnaire would apply scales and instruments to measure trust and its different dimensions. These are evident in some areas such as in surveys of patient trust in a medical context [18]. However, in the specific area of researcher trust in data, there is little previous research. Hence, focus group discussions could be used in a first phase for the construction, testing, and refinement of relevant and meaningful survey questions. In addition to providing relevant data for analysis, the survey could also serve as a sampling frame to identify interviewees who report varying levels of trust for semi-structured interviews that further explore the reasons for variations in levels of trust.

Once, we understand the criteria by which life scientists assess data quality, we can think about adapting the incentives in our research systems and culture in a way that increases data robustness and, in turn, advances progress and enhances the return-of-investment of research funding. More evidence on concrete questionable research practices that are anticipated to affect reproducibility [2] will enable the design of tailored measures that can then be implemented and systematically monitored for effectiveness.

Taken together, we feel that meta-research using established sociological and philosophical expertise and methodologies to study researcher trust in scientific data is a straightforward approach to establish a better understanding of the extent of the reproducibility crisis and to tackle the underlying factors.

CRediT authorship contribution statement

Calnan Michael: Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Kirchin Simon:** Conceptualization, Methodology, Writing – review & editing. **Michaelis Martin:** Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Roberts David L:** Conceptualization, Methodology, Writing – review & editing. **Wass Mark N:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

None.

References

- 1 M.N. Wass, L. Ray, M. Michaelis, Understanding of researcher behavior is required to improve data reliability, *Gigascience* 8 (5) (2019), giz017, <https://doi.org/10.1093/gigascience/giz017>.
- 2 J. de Ridder, How to trust a scientist, *Stud. Hist. Philos. Sci.* 93 (2022) 11–20, <https://doi.org/10.1016/j.shpsa.2022.02.003>.
- 3 C. Bernard, Stop reproducing the reproducibility crisis, *eNeuro* 10 (2) (2023), ENEURO.0032-23.2023, <https://doi.org/10.1523/ENEURO.0032-23.2023>.
- 4 A. Mullard, Half of top cancer studies fail high-profile reproducibility effort, *Nature* 600 (7889) (2021) 368–369, <https://doi.org/10.1038/d41586-021-03691-0>.
- 5 T.M. Errington, M. Mathur, C.K. Soderberg, A. Denis, N. Perfito, E. Iorns, B.A. Nosek, Investigating the replicability of preclinical cancer biology, *Elife* 10 (2021), e71601, <https://doi.org/10.7554/eLife.71601>.
- 6 D. Fanelli, R. Costas, J.P. Ioannidis, Meta-assessment of bias in science, *Proc. Natl. Acad. Sci. USA* 114 (14) (2017) 3714–3719, <https://doi.org/10.1073/pnas.1618569114>.
- 7 B. Voelkl, H. Würbel, A reaction norm perspective on reproducibility, *Theory Biosci.* 140 (2) (2021) 169–176, <https://doi.org/10.1007/s12064-021-00340-y>.
- 8 M. Wadman, FDA no longer has to require animal testing for new drugs, *Science* 379 (6628) (2023) 127–128, <https://doi.org/10.1126/science.adg6276>.
- 9 I.G. Reddin, T.R. Fenton, M.N. Wass, M. Michaelis, Large inherent variability in data derived from highly standardised cell culture experiments, *Pharm. Res* 188 (2023), 106671, <https://doi.org/10.1016/j.phrs.2023.106671>.
- 10 A. Dreber, T. Pfeiffer, J. Almenberg, S. Isaksson, B. Wilson, Y. Chen, B.A. Nosek, M. Johannesson, Using prediction markets to estimate the reproducibility of scientific research, *Proc. Natl. Acad. Sci. USA* 112 (50) (2015) 15343–15347, <https://doi.org/10.1073/pnas.1516179112>.

- 11 M. Gordon, D. Viganola, A. Dreber, M. Johannesson, T. Pfeiffer, Predicting replicability-Analysis of survey and prediction market data from large-scale forecasting projects, *PLoS One* 16 (4) (2021), e0248780, <https://doi.org/10.1371/journal.pone.0248780>.
- 12 B. Latour, S. Woolgar, *Laboratory life: the construction of scientific facts*, Princeton University Press, 1986. ISBN 9781400820412.
- 13 M.W. Calnan, R. Rowe, *Trust Matters In Health Care*, Open University Press, 2008, p. 224 pp. ISBN 978-0-335-22283-4.
- 14 K.E. Hubbard, S.D. Dunbar, Perceptions of scientific research literature and strategies for reading papers depend on academic career stage, *PLoS One* 12 (12) (2017), e0189753, <https://doi.org/10.1371/journal.pone.0189753>.
- 15 B. Maher, Cancer reproducibility project scales back ambitions, *Nature* (2015), <https://doi.org/10.1038/nature.2015.18938>.
- 16 M.S. Hagger, Developing an open science 'mindset', *Health Psychol. Behav. Med* 10 (1) (2021) 1–21, <https://doi.org/10.1080/21642850.2021.2012474>.
- 17 M. Calnan, Quantitative Survey Methods in Health Research, in: Saks, Allsop (Eds.), *Researching Health: Qualitative, Quantitative and Mixed Methods*, SAGE Publications, London, 2018. Chapter 13.
- 18 S. Aboueid, H. Herati, I. Maria, G. Nascimento, P. Ward, P. Brown, M. Calnan, C. Perlman, S. Meyer, How do you measure trust in social institutions and health professionals? A systematic review of the literature (2012–2021), *Sociol. Compass* 17 (2023), e13101, <https://doi.org/10.1111/soc4.13101>.