

Researcher conduct determines data reliability

Mark N. Wass^{1*}, Larry Ray², Martin Michaelis^{1*}

¹ Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, UK

² School of Social Policy, Sociology and Social Research, University of Kent, Canterbury, UK

E-mail addresses: Mark N. Wass, M.N.Wass@kent.ac.uk; Larry Ray, L.J.Ray@kent.ac.uk; Martin Michaelis, M.Michaelis@kent.ac.uk

*Correspondence to: Mark N. Wass, M.N.Wass@kent.ac.uk; Martin Michaelis, M.Michaelis@kent.ac.uk

Abstract

A lack of data reproducibility (“reproducibility crisis”) is debated across many scientific and medical disciplines. A systematic analysis of the available evidence on the reliability of scientific data revealed that, although the existence of a reproducibility crisis is widely perceived, conclusive data on the scale of the problem are lacking. Most importantly we found that, although the debate is largely focused on methodological issues, researcher conduct defines research standards and in turn data reliability. The availability of reliable methods itself does not guarantee good practice. Moreover, research is typically characterised by a lack of established methods due to its immanent novelty. Despite the crucial importance of researcher conduct, research and conclusive data on the determinants of researcher behaviour are missing. In conclusion, meta-research is urgently needed that establishes an understanding of the factors that determine researcher behaviour. This knowledge can then be used to implement and iteratively improve measures, which incentivise researchers to apply the highest standards resulting in high quality data.

Key words: reproducibility crisis, replication crisis, data reliability, bias, publication bias, meta-research

Background

A lack of data reproducibility ("reproducibility crisis") is debated across many medical and scientific disciplines [1-12]. It seems to receive increasing attention as demonstrated by the rise in articles indexed in PubMed [13] related to the terms "reproducibility crisis" and "replication crisis" (Figure 1). This finding is in accordance with another recent analysis that indicated a rapidly increasing number of scientific articles within a "crisis narrative" [14]. Factors suggested to affect reproducibility include (a lack of) methodological standards, (unconscious) bias, pressure related to the need to attract grants and publish in high impact journals, and publication bias favouring the publication of novel ("positive") findings and discouraging the publication of confirmatory findings and "negative" results [3,11,15-22]. Some authors argue that a high proportion (up to 90%) of research money is wasted [2-7], although this very pessimistic view is not unambiguously shared [14]. Nevertheless, the perception of a reproducibility crisis seems to be common among researchers. In a *Nature* survey, 52% of the 1576 respondents from all science areas agreed that a reproducibility crisis exists [23].

Method

To determine what is known about the extent and causes of the reproducibility crisis and how the scientific community deal with it, we performed a systematic analysis of the available literature on the reproducibility of research findings. Literature searches were performed using PubMed on 3rd November 2017 using the search terms "replication crisis" (resulting in 260 hits) and "reproducibility crisis" (306 hits). Cell line misidentification has been discussed as a significant cause of unreliable results since first mammalian cell lines were established [24,25]. Thus, we also used the search term "cell line authentication" (17th November 2017, 147 hits). We identified a further

104 relevant articles from the analysed articles. After removing the overlaps, 791 articles were analysed resulting in 280 relevant articles in English (Suppl. Table 1; Suppl. Figure 1).

Results

Scale of crisis remains unclear

Despite the high visibility of the issue, systematic research and in turn conclusive evidence on the scale of a potential reproducibility crisis is lacking. In a survey among faculty and trainees at the MD Anderson Cancer Center, about 50% of the participants reported that they had failed to reproduce published data at least once [26]. Similarly, in a *Nature* survey >70% of the 1576 respondents stated that they had been unable to reproduce data at least once [23]. However, systematic data that would enable the reliable quantification of the issue was not identified. We found 17 direct replication attempts among the articles that focused on methodology (Suppl. Table 1). This included nine sciences, six psychology studies, one public health study, and one economics study.

The scientific studies were part of the “Reproducibility Project: Cancer Biology” by the Center for Open Science [27] and Science Exchange [28]. In this project, findings from 29 high-profile scientific publications will be independently replicated [29,30,31]. So far, the results of nine replication studies have been reported. Important parts of the original paper could be reproduced in four studies [32-35]. The results from two replication studies could not be interpreted [36,37], and one study failed to replicate the original findings [38]. In the two most recent reports, parts of the original studies were reproduced while other parts were not [39,40] (Suppl. Table 1).

Psychological studies reported varying levels of successful replication. Two psychological studies suggested very low levels of reproducibility [41,42], and a study

by the Open Science Collaboration reported the successful replication of 39 of 100 psychological studies [9]. However, further studies replicated a majority of the analysed effects [43] or confirmed previous findings [44,45]. Finally, there were a successful confirmation study from the public health area [46] and a partially successfully economics replication study [47] (Suppl. Table 1).

In the clinical research field, an analysis of follow-up publications of 49 original clinical research studies, which had been published between 1990-2003 and had each acquired more than 1000 citations, revealed that seven (16%) were not confirmed by subsequent studies, seven (16%) had reported stronger effects than those found in subsequent studies, 20 (44%) were successfully replicated, and for 11 (24%) follow-up data was not available [1].

Reproducibility data was also provided in articles other than original research articles. Researchers from drug companies reported that only six out of 53 studies (11%) [5] or 16 out of 67 studies (24%) [3] had been successfully reproduced. However, these data were published as a Comment [5] and a Correspondence [3] without presentation of detailed data. Hence, the exact nature of the investigations and the criteria for reproducibility remain elusive.

Taken together these findings show that despite anecdotal reports of data irreproducibility, the scale of the issue remains unclear. Most replication attempts focus on highly cited early-stage studies. This may not adequately reflect the general reproducibility of research findings. A meta-assessment of bias in the sciences observed a significant risk of small, early, and highly cited studies to overestimate effects [48]. Further, failed and successful replication attempts would need to be systematically analysed together to provide meaningful insights. However, such

studies are not available. A psychology study estimated that only about 1% of studies are subject to replication attempts [49].

The determination of the scale of the problem may be further complicated by the absence of clear criteria that define the successful or unsuccessful repetition of a study. Ten articles represented a dispute on the consistency of data from two large pharmacogenomics screens in cancer cell lines [50,51]. Six of these articles reported discrepancies between the datasets, while four reported consistency. All six articles that reported discrepancies were published by the same research group, whereas the articles reporting consistency were published by four different research groups (Suppl. Table 1). However, the dispute does not appear to have been resolved. In this context, a modelling study from the psychology field suggests that the criteria for reproducibility may sometimes be interpreted in an unrealistically strict fashion [52].

Researcher conduct defines data quality

242 of the 280 relevant articles considered methodological shortcomings. Other potential issues were less commonly addressed and included research culture and incentives (27 times), publication bias (20), and transparency/ availability of raw data (14) (Figure 2, Suppl. Table 1).

126 of the 280 articles contained original data directly related to reproducibility (Figure 2, Suppl. Table 1). Again, the vast majority reported on methodological issues (105 articles). A few (12) reported on publication bias (Figure 2, Suppl. Table 1). Together, these data indicate a strong focus on methodological issues in the debate about data reproducibility.

The issue of transparency/ availability of raw data was mentioned in total 14 times (Figure 2, Suppl. Table 1) and has been recognised by research funders and scientific journals [53]. For example, the UK funders Medical Research Council, Academy of

Medical Sciences, Wellcome Trust, and Biotechnology and Biological Sciences Research Council published a common report on data reproducibility [54] and the World Economic Forum set up a “Code of Ethics for Researchers” [55]. Journals from publishers including the Nature Publishing group and EMBO Press introduced 'publication checklists' [see e.g. 56,57]. Nature also published a special collection on reproducibility in 2013 [58]. However, data on the effects of increased transparency are lacking. A survey reported psychologists to be open to changes to data collection, reporting, and publication practices, but to be less positive about mandatory conditions of publication [59]. One preprint posted on bioRxiv suggested that the introduction of a checklist by Nature had indeed improved study design and the transparency of data [60], but data indicating whether this translated into improved reproducibility is not yet available.

Despite the focus of the debate on research methodology, it remains unclear whether (and if yes, to which extent) a lack of reproducibility may be caused by a lack of (knowledge of) appropriate methods and to which extent the significance of data can be improved by standardisation. In animal experiments, batch-to-batch variation was described even under highly standardised conditions in the same lab [61]. Experiment heterogenisation and multi-laboratory design was suggested to produce more reliable data [61-65]. In the area of drug discovery, clear requirements for the generation of reproducible data have been suggested [see e.g. 19,21,22,66]. However, data on the implementation of such measures and their efficacy with regard to improved reproducibility are not available.

With regard to the use of appropriate methodologies, cell line misidentification has been an area of concern since the first cell lines were established [24,25]. 49 of the 105 articles, which reported original data on methodological issues, focused on cell

line authentication (Figure 2, Suppl. Table 1). Although short tandem repeat (STR) analysis has been available and promoted as a reliable authentication method since at least 2001 [67], very recent articles still demonstrate that the use of misidentified cell lines remains an issue [68,69]. Similarly, a meta-analysis considering 60-years of articles indicated that the statistical power of behavioural sciences studies has not increased, although the need to increase the statistical power was repeatedly discussed and demonstrated [70]. Hence, the availability of suitable and reliable methods is not sufficient to guarantee their appropriate and consequent use. Additionally, it is a characteristic of research that experiments are performed and methodologies are used for the first time. Consequently, researcher conduct and the research culture are critical to ensure the highest possible reliability of data. Hence, more focus and effort needs to be invested to understand how researchers report and present their data.

Researcher careers are driven by publications in as highly prestigious research journals as possible to gain visibility and attract research funding [19,71]. This requires the presentation of novel, significant findings, which incentivises the publication of 'positive' findings and discourages the publication of 'negative' findings. This may also incentivise smaller (potentially underpowered) studies, because they are more likely to produce significant results than larger studies [19,71]. A modelling study indicated that the best strategy to produce significant findings and optimise research output is to perform small studies that only have 10-40% statistical power, which would result in half of the studies reporting false-positive findings [72]. Further modelling studies suggested that a focus on novel findings and positive results undermines the rigorousness of science [70,73]. Accordingly, early, highly-cited studies seem to be more likely to present exaggerated findings [48].

Contribution of publication bias

A focus on 'positive' results also favours 'publication bias', i.e. 'positive' results are more likely to be published than 'negative' findings. Hence, the available literature does not appropriately represent the totality of experiments that have been performed, because many 'negative' results remain unpublished ("file drawer problem"). Additionally, 'positive' findings are more likely to be published in prestigious journals than 'negative' findings [18,19,74].

We identified twelve articles that provided data on publication bias (Suppl. Table 1). One study reported the overestimation of the importance of anticipated prognostic factors in various types of cancer due to publication bias [75]. A follow-up study, which investigated 1,915 research articles on prognostic markers in cancer, found that >90% of studies reported positive prognostic correlations [76]. Less than 1.5% of the investigated articles provided purely 'negative' data. If 'negative' findings were presented, this typically happened in the context of other significant correlations ('positive' findings), or the authors followed up on non-significant trends and tried to defend the importance of the investigated markers despite the lack of significance [76]. This illustrates that negative results are not commonly published. The evaluation of meta-analyses on cancer biomarkers and the analysis of animal studies on stroke and neurological diseases also suggested a bias towards the publication of 'positive results' [77-79].

Further, a similar publication bias was reported for both clinical trials [80] and psychological studies [81,82]. Accordingly, a large analysis using US data concluded that there is a general publication bias towards the publication of 'positive' results across the academic disciplines [83]. This bias seems to be more pronounced, the less results are characterised by exact quantitative data [84]. Notably, this topic

becomes complicated by findings that suggest that meta-research on publication bias may itself be subject to publication bias [85]. Taken together, there is convincing evidence that a bias favouring the publication of 'positive' findings exists and that it can affect the reliability of publicly available data, although the scale of the impact is not clear.

Determinants of researcher conduct are unclear

Researcher conduct defines the reliability of findings also beyond publication bias. This is highly relevant as original research is typically defined by a significant level of novelty in the absence of established standards. Findings are often made using novel (combinations of) approaches together with (novel) model systems and/ or (novel) data for the first time, i.e. before tested and standardised approaches are available. It is fair to assume that the incentives provided in a research environment substantially influence researcher behaviour. One substantial meta-analysis based on data from 18 surveys concluded that a pooled weighted estimate of 1.97% (crude unweighted mean: 2.59%) of the respondents admitted to have fabricated, falsified or modified data or results at least once. 14.12% (crude unweighted mean: 16.66%) reported to personally know of a colleague who had done so [86]. Hence, there is evidence of questionable research practices, but the actual extent, the influence of the research environment and its incentives, and the concrete effect on data reliability remain elusive.

To identify studies that investigated researcher (mis)conduct in response to the pressures and incentives of the research environment, we performed an additional literature search using the term "misconduct" (although not all forms of scientific misconduct affect data reproducibility, e.g. plagiarism, gift authorship, and circumventing animal-subjects research requirements) in PubMed (18th January

2018). This search resulted in 9394 hits (Suppl. Table 2). We identified three relevant articles and based on these articles three additional ones (Suppl. Table 2). These articles supported the idea that the pressure to publish contributes to questionable research practices that may affect the reproducibility of findings.

A survey analysing the answers of 3247 early- and mid-career scientists suggested that a feeling of injustice may contribute to questionable research practices, which may affect reproducibility [87,88]. Focus group discussions involving 51 scientists from research universities revealed that the pressure to produce outputs also promotes questionable research practices [89], which may affect reproducibility. In a survey among 315 Flemish biomedical scientists, 15% of the respondents admitted that they had fabricated, falsified, plagiarised, or manipulated data in the past 3 years. 72% rated the publication pressure as "too high" [90]. A follow-up qualitative focus group interview study among Dutch biomedical researchers suggested that the current publication culture leads to questionable research practices among junior and senior biomedical scientists [91]. Hence, there is some initial evidence that the pressure associated with a highly competitive environment affects researcher conduct, which in turn affects the reliability and reproducibility of data. Again, however, the actual scale remains elusive.

Conclusions

Our findings demonstrate the need for systematic meta-research on the issue of data reproducibility. A reproducibility crisis is widely recognised among researchers from many different fields. There is no shortage of suggestions on how data reproducibility could be improved, but quantitative data on the subject (including the scale of the problem) are largely missing. Currently, there is a strong focus on methodology. However, ongoing issues with the use of misidentified cell lines illustrate that problems may persist, even despite effective standards being available. Further, it is in the nature of research to do things for the first time before established methods become available. Hence, research standards are primarily defined by the conduct of researchers and their rigour in the acquisition, analysis, and interpretation of data. There are some indications that the high pressure associated with a competitive environment increases the preparedness of researchers to lower their ethical standards, but the available information remains scarce. Consequently, a major focus of meta-research on data reproducibility will need to be put on researcher behaviour, especially in relation to “boundary work”, that is, the ways researchers draw the boundaries between the permissible and the non-permissible [92]. Only measures that are based on a detailed understanding of researcher behaviour and that are closely monitored for efficacy (and iteratively improved) will make it possible to amend our research system in a way that it provides the right incentives to ensure that researchers apply the highest possible standards and provide high quality data.

References

- 1) Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294:218-28.
- 2) Young SS, Bang H, Oktay K. Cereal-induced gender selection? Most likely a multiple testing false positive. *Proc Biol Sci*. 2009;276:1211–2; discussion 1213.
- 3) Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712.
- 4) Young SS, Karr A. Deming, data and observational studies: a process out of control and needing fixing. *Significance*. 2011;9:122-6.
- 5) Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483:531-3.
- 6) Peers IS, Ceuppens PR, Harbron C. In search of preclinical robustness. *Nat Rev Drug Discov*. 2012;11:733–4.
- 7) Young SS, Miller HI. Are medical articles true on health, disease? Sadly, not as often as you might think. *Genetic Engineering and Biotechnology News* 2014;34:7-9.
- 8) Begley CG, Buchan AM, Dirnagl U. Robust research: Institutions must do their part for reproducibility. *Nature*. 2015;525:25-7.
- 9) Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015;349:aac4716.
- 10) Kousta S, Ferguson C, Ganley E. Meta-Research: Broadening the Scope of PLOS Biology. *PLoS Biol*. 2016;14:e1002334.
- 11) Lilienfeld SO. Psychology's Replication Crisis and the Grant Culture: Righting the Ship. *Perspect Psychol Sci*. 2017;12:660-4.
- 12) Hutson M. Artificial intelligence faces reproducibility crisis. *Science*. 2018;359:725-6.

- 296 13) <https://www.ncbi.nlm.nih.gov/pubmed>. Accessed 12 January 2018.
- 297 14) Fanelli D. Opinion: Is science really facing a reproducibility crisis, and do we need
298 it to? Proc Natl Acad Sci U S A. 2018;115:2628-2631.
- 299 15) Casadevall A, Fang FC. Reforming science: methodological and cultural reforms.
300 Infect Immun. 2012;80:891-6.
- 301 16) Fang FC, Casadevall A. Reforming science: structural reforms. Infect Immun.
302 2012;80:897-901.
- 303 17) Ioannidis JP. How to make more published research true. PLoS Med.
304 2014;11:e1001747.
- 305 18) Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz
306 KF, Tibshirani R. Increasing value and reducing waste in research design, conduct,
307 and analysis. Lancet. 2014;383:166-75.
- 308 19) Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for
309 basic and preclinical research. Circ Res. 2015;116:116-26.
- 310 20) Jarvis MF, Williams M. Irreproducibility in Preclinical Biomedical Research:
311 Perceptions, Uncertainties, and Knowledge Gaps. Trends Pharmacol Sci.
312 2016;37:290-302.
- 313 21) Kaelin WG Jr. Publish houses of brick, not mansions of straw. Nature.
314 2017;545:387.
- 315 22) Kaelin WG Jr. Common pitfalls in preclinical cancer target validation. Nat Rev
316 Cancer. 2017;17:425-40.
- 317 23) Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533:452-4.
- 318 24) American Type Culture Collection Standards Development Organization
319 Workgroup ASN-0002. Cell line misidentification: the beginning of the end. Nat Rev
320 Cancer. 2010;10:441-8.

- 321 25) Capes-Davis A, Neve RM. Authentication: A Standard Problem or a Problem of
322 Standards? PLoS Biol. 2016;14:e1002477.
- 323 26) Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A survey on data
324 reproducibility in cancer research provides insights into our limited ability to translate
325 findings from the laboratory to the clinic. PLoS One. 2013;8(5):e63221.
- 326 27) <https://cos.io>. Accessed on 7 March 2018.
- 327 28) <https://www.scienceexchange.com>. Accessed on 7 March 2018.
- 328 29) Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open
329 investigation of the reproducibility of cancer biology research. Elife. 2014 Dec 10;3.
330 doi: 10.7554/eLife.04333.
- 331 30) Baker M, Dolgin E. Cancer reproducibility project releases first results. Nature.
332 2017;541:269-270.
- 333 31) [https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-](https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology)
334 [biology](https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology). Accessed on 7 March 2018.
- 335 32) Aird F, Kandela I, Mantis C; Reproducibility Project: Cancer Biology. Replication
336 Study: BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Elife.
337 2017;6. pii: e21253.
- 338 33) Kandela I, Aird F; Reproducibility Project: Cancer Biology. Replication Study:
339 Discovery and preclinical validation of drug indications using compendia of public gene
340 expression data. Elife. 2017;6. pii: e17044.
- 341 34) Shan X, Fung JJ, Kosaka A, Danet-Desnoyers G; Reproducibility Project: Cancer
342 Biology. Replication Study: Inhibition of BET recruitment to chromatin as an effective
343 treatment for MLL-fusion leukaemia. Elife. 2017;6. pii: e25306.
- 344 35) Showalter MR, Hatakeyama J, Cajka T, VanderVorst K, Carraway KL, Fiehn O;
345 Reproducibility Project: Cancer Biology. Replication Study: The common feature of

346 leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity
347 converting alpha-ketoglutarate to 2-hydroxyglutarate. *Elife*. 2017 Jun 27;6. pii:
348 e26030.

349 36) Horrigan SK; Reproducibility Project: Cancer Biology. Replication Study: The
350 CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for
351 human solid tumors. *Elife*. 2017;6. pii: e18173.

352 37) Horrigan SK, Courville P, Sampey D, Zhou F, Cai S; Reproducibility Project:
353 Cancer Biology. Replication Study: Melanoma genome sequencing reveals frequent
354 PREX2 mutations. *Elife*. 2017;6. pii: e21634. doi: 10.7554/eLife.21634.

355 38) Mantis C, Kandela I, Aird F; Reproducibility Project: Cancer Biology. Replication
356 Study: Coadministration of a tumor-penetrating peptide enhances the efficacy of
357 cancer drugs. *Elife*. 2017;6. pii: e17584. doi: 10.7554/eLife.17584.

358 39) Lewis LM, Edwards MC, Meyers ZR, Talbot CC, Hao H, Blum D, et al. Replication
359 Study: Transcriptional amplification in tumor cells with elevated c-Myc. *Elife*. 2018;7.
360 pii: e30274.

361 40) Vanden Heuvel JP, Maddox E, Maalouf SW; Reproducibility Project: Cancer
362 Biology, Iorns E, et al. Replication Study: Systematic identification of genomic markers
363 of drug sensitivity in cancer cells. *Elife*. 2018;7. pii: e29747.

364 41) Boekel W, Wagenmakers EJ, Belay L, Verhagen J, Brown S, Forstmann BU. A
365 purely confirmatory replication study of structural brain-behavior correlations. *Cortex*.
366 2015;66:115-33.

367 42) Emmerling F, Martijn C, Alberts HJ, Thomson AC, David B, Kessler D, et al. The
368 (non-)replicability of regulatory resource depletion: A field report employing non-
369 invasive brain stimulation. *PLoS One*. 2017;12:e0174331.

- 370 43) Klein RA, Ratliff KA, Vianello M, Adams RB Jr, Bahník Š, Bernstein MJ, et al.
 371 Investigating variation in replicability: A “many labs” replication project. *Soc Psychol.*
 372 2014;45:142–52.
- 373 44) Ahmad MM. Psychometric evaluation of the Cognitive Appraisal of Health Scale
 374 with patients with prostate cancer. *J Adv Nurs.* 2005;49:78-86.
- 375 45) Zwaan RA, Pecher D, Paolacci G, Bouwmeester S, Verkoeijen P, Dijkstra K, et al.
 376 Participant Nonnaiveté and the reproducibility of cognitive psychology. *Psychon Bull*
 377 *Rev.* 2017 Jul 25. doi: 10.3758/s13423-017-1348-y.
- 378 46) Firn M, Alonso-Vicente M, Hubbeling D, Holley J, White SJ, Jones B. The
 379 replacement of assertive outreach services by reinforcing local community teams: a
 380 replication study reporting comparative outcomes and patient reported experience. *J*
 381 *Ment Health.* 2017;27:1-7.
- 382 47) Camerer CF, Dreber A, Forsell E, Ho TH, Huber J, Johannesson M, et al.
 383 Evaluating replicability of laboratory experiments in economics. *Science.*
 384 2016;351:1433-6.
- 385 48) Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. *Proc Natl*
 386 *Acad Sci U S A.* 2017;114:3714-9.
- 387 49) Makel MC, Plucker JA, Hegarty B. Replications in Psychology Research: How
 388 Often Do They Really Occur? *Perspect Psychol Sci.* 2012;7:537-42.
- 389 50) Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al.
 390 The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug
 391 sensitivity. *Nature.* 2012;483:603-7.
- 392 51) Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al.
 393 Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature.*
 394 2012;483:570-5.

- 395 52) Stanley DJ, Spence JR. Expectations for Replications: Are Yours Realistic?
396 Perspect Psychol Sci. 2014;9:305-18.
- 397 53) Nature Editorial. A code of ethics to get scientists talking. Nature. 2018;555:5.
- 398 54) <https://acmedsci.ac.uk/download?f=file&i=32558>. Accessed on 7 March 2018.
- 399 55) <http://widgets.weforum.org/coe/>. Accessed on 7 March 2018.
- 400 56) Nature Announcement. Reducing our irreproducibility. Nature. 2013;496:398.
- 401 57) Nature Editorial. Steps towards transparency in research publishing. Nature.
402 2017;549:431.
- 403 58) <https://www.nature.com/collections/prbfkwmwvz/>. Accessed on 7 March 2018.
- 404 59) Fuchs HM, Jenny M, Fiedler S. Psychologists Are Open to Change, yet Wary of
405 Rules. Perspect Psychol Sci. 2012;7:639-42.
- 406 60) Macleod MR, The NPQIP Collaborative group. Findings of a retrospective,
407 controlled cohort study of the impact of a change in Nature journals' editorial policy for
408 life sciences research on the completeness of reporting study design and execution.
409 bioRxiv 2017. doi: <https://doi.org/10.1101/187245>
- 410 61) Karp NA. Reproducible preclinical research-Is embracing variability the answer?
411 PLoS Biol. 2018;16:e2005413.
- 412 62) Karp NA, Speak AO, White JK, Adams DJ, Hrabé de Angelis M, Hérault Y, et al.
413 Impact of temporal variation on design and analysis of mouse knockout phenotyping
414 studies. PLoS One. 2014;9:e111239.
- 415 63) Kafkafi N, Golani I, Jaljuli I, Morgan H, Sarig T, Würbel H, et al. Addressing
416 reproducibility in single-laboratory phenotyping experiments. Nat Methods.
417 2017;14:462-4.
- 418 64) Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research
419 improves with heterogeneity of study samples. PLoS Biol. 2018;16:e2003693.

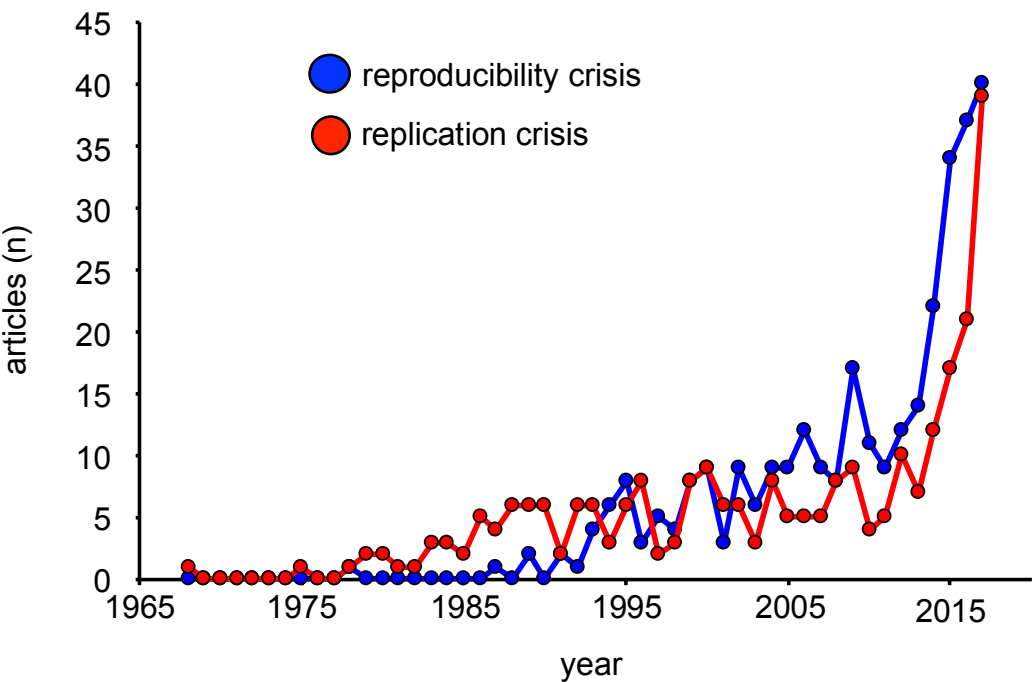
- 420 65) Milcu A, Puga-Freitas R, Ellison AM, Blouin M, Scheu S, Freschet GT, et al.
 421 Genotypic variability enhances the reproducibility of an ecological study. *Nat Ecol*
 422 *Evol.* 2018;2:279-2.
- 423 66) Begley CG. Six red flags for suspect work. *Nature* 2013;497:433-4.
- 424 67) Masters JR, Thomson JA, Daly-Burns B, Reid YA, Dirks WG, Packer P, et al. Short
 425 tandem repeat profiling provides an international reference standard for human cell
 426 lines. *Proc Natl Acad Sci U S A.* 2001;98:8012-7.
- 427 68) Vaughan L, Glänzel W, Korch C, Capes-Davis A. Widespread Use of Misidentified
 428 Cell Line KB (HeLa): Incorrect Attribution and Its Impact Revealed through Mining the
 429 Scientific Literature. *Cancer Res.* 2017;77:2784-8.
- 430 69) Wang M, Yang M, Liu Y, Huang Y, Ye F, Zheng C, Shen C. Investigation of cross-
 431 contamination among human cell lines used in China. *Int J Cancer.* 2017 Aug 10. doi:
 432 10.1002/ijc.30923.
- 433 70) Smaldino PE, McElreath R. The natural selection of bad science. *R Soc Open Sci.*
 434 2016;3:160384.
- 435 71) Brembs B. Prestigious Science Journals Struggle to Reach Even Average
 436 Reliability. *Front Hum Neurosci.* 2018;12:37.
- 437 72) Higginson AD, Munafò MR. Current Incentives for Scientists Lead to
 438 Underpowered Studies with Erroneous Conclusions. *PLoS Biol.* 2016;14:e2000995.
- 439 73) Grimes DR, Bauch CT, Ioannidis JPA. Modelling science trustworthiness under
 440 publish or perish pressure. *R Soc Open Sci.* 2018;5:171511.
- 441 74) Nissen SB, Magidson T, Gross K, Bergstrom CT. Publication bias and the
 442 canonization of false facts. *Elife.* 2016;5. pii: e21451.
- 443 75) Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic
 444 factor studies. *J Natl Cancer Inst.* 2005;97:1043-55.

- 445 76) Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic
446 markers report statistically significant results. *Eur J Cancer*. 2007;43:2559-79.
- 447 77) Tsilidis KK, Papatheodorou SI, Evangelou E, Ioannidis JP. Evaluation of excess
448 statistical significance in meta-analyses of 98 biomarker associations with cancer risk.
449 *J Natl Cancer Inst*. 2012;104:1867-78.
- 450 78) Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias
451 in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*.
452 2010;8:e1000344.
- 453 79) Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et
454 al. Evaluation of excess significance bias in animal studies of neurological diseases.
455 *PLoS Biol*. 2013;11:e1001609.
- 456 80) Hall R, de Antueno C, Webber A; Canadian Research Ethics Board. Publication
457 bias in the medical literature: a review by a Canadian Research Ethics Board. *Can J*
458 *Anaesth*. 2007;54:380-8.
- 459 81) Bakker M, van Dijk A, Wicherts JM. The Rules of the Game Called Psychological
460 Science. *Perspect Psychol Sci*. 2012;7:543-54.
- 461 82) Ferguson CJ, Heene M. A Vast Graveyard of Undead Theories: Publication Bias
462 and Psychological Science's Aversion to the Null. *Perspect Psychol Sci*. 2012;7:555-
463 61.
- 464 83) Fanelli D. Do pressures to publish increase scientists' bias? An empirical support
465 from US States Data. *PLoS One*. 2010;5:e10271.
- 466 84) Fanelli D. "Positive" results increase down the Hierarchy of the Sciences. *PLoS*
467 *One*. 2010;5:e10068.
- 468 85) Dubben HH, Beck-Bornholdt HP. Systematic review of publication bias in studies
469 on publication bias. *BMJ*. 2005;331:433-4.

- 470 86) Fanelli D. How many scientists fabricate and falsify research? A systematic review
471 and meta-analysis of survey data. PLoS One. 2009;4:e5738.
- 472 87) Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature.
473 2005;435:737-8.
- 474 88) Martinson BC, Anderson MS, Crain AL, de Vries R. Scientists' perceptions of
475 organizational justice and self-reported misbehaviors. J Empir Res Hum Res Ethics.
476 2006;1:51-66.
- 477 89) de Vries R, Anderson MS, Martinson BC. Normal Misbehavior: Scientists Talk
478 about the Ethics of Research. J Empir Res Hum Res Ethics. 2006;1:43-50.
- 479 90) Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific
480 misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9:64-71.
- 481 91) Tijdink JK, Schipper K, Bouter LM, Maclaine Pont P, de Jonge J, Smulders YM.
482 How do scientists perceive the current publication culture? A qualitative focus group
483 interview study among Dutch biomedical researchers. BMJ Open. 2016;6:e008681.
- 484 92) Hesselmann F, Wienefoet V, Reinhart M. Measuring Scientific Misconduct—
485 Lessons from Criminology. Publications 2014;2(3):61-70.

486 **Figures**

Figure 1



487

488

489 **Figure 1.** Number of articles that are identified by the search terms “replication crisis”
490 (red) or “reproducibility crisis” (blue) per year from 1965 to 2017 in PubMed
491 (www.ncbi.nlm.nih.gov/pubmed, data accessed on 12th January 2018).

492

Figure 2

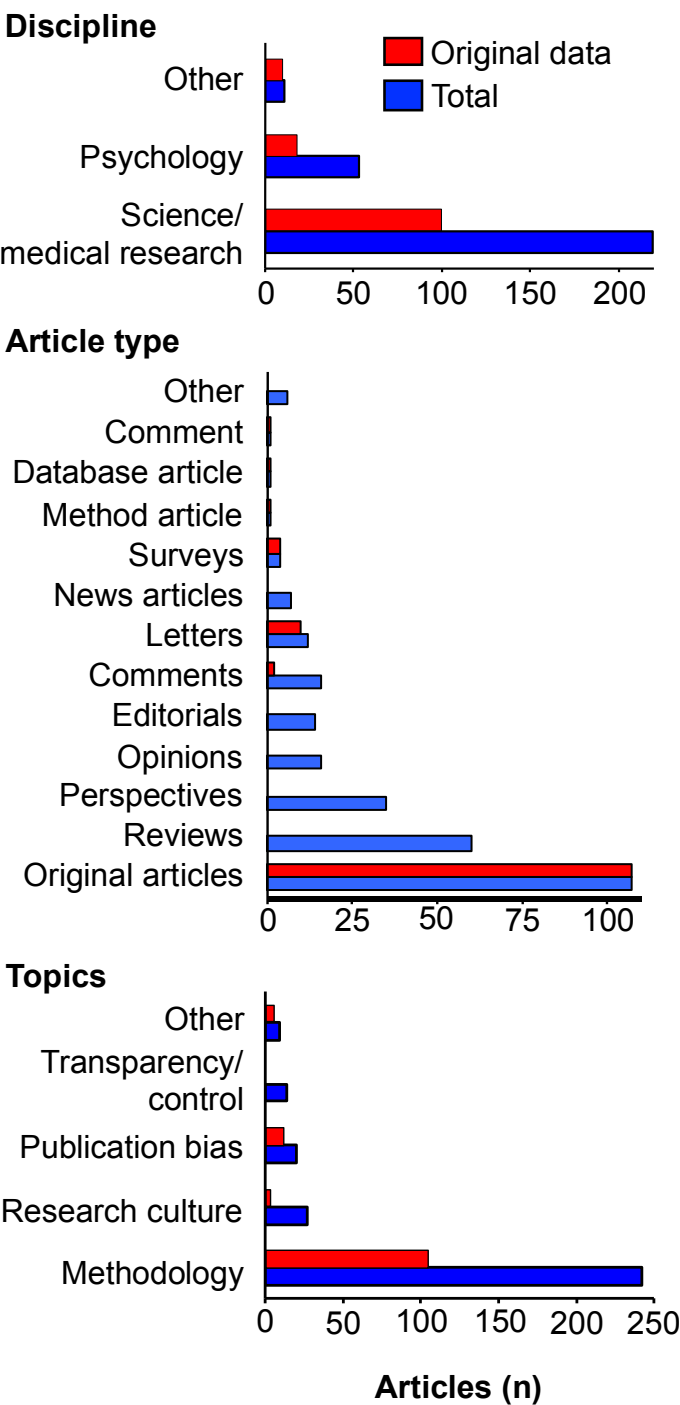
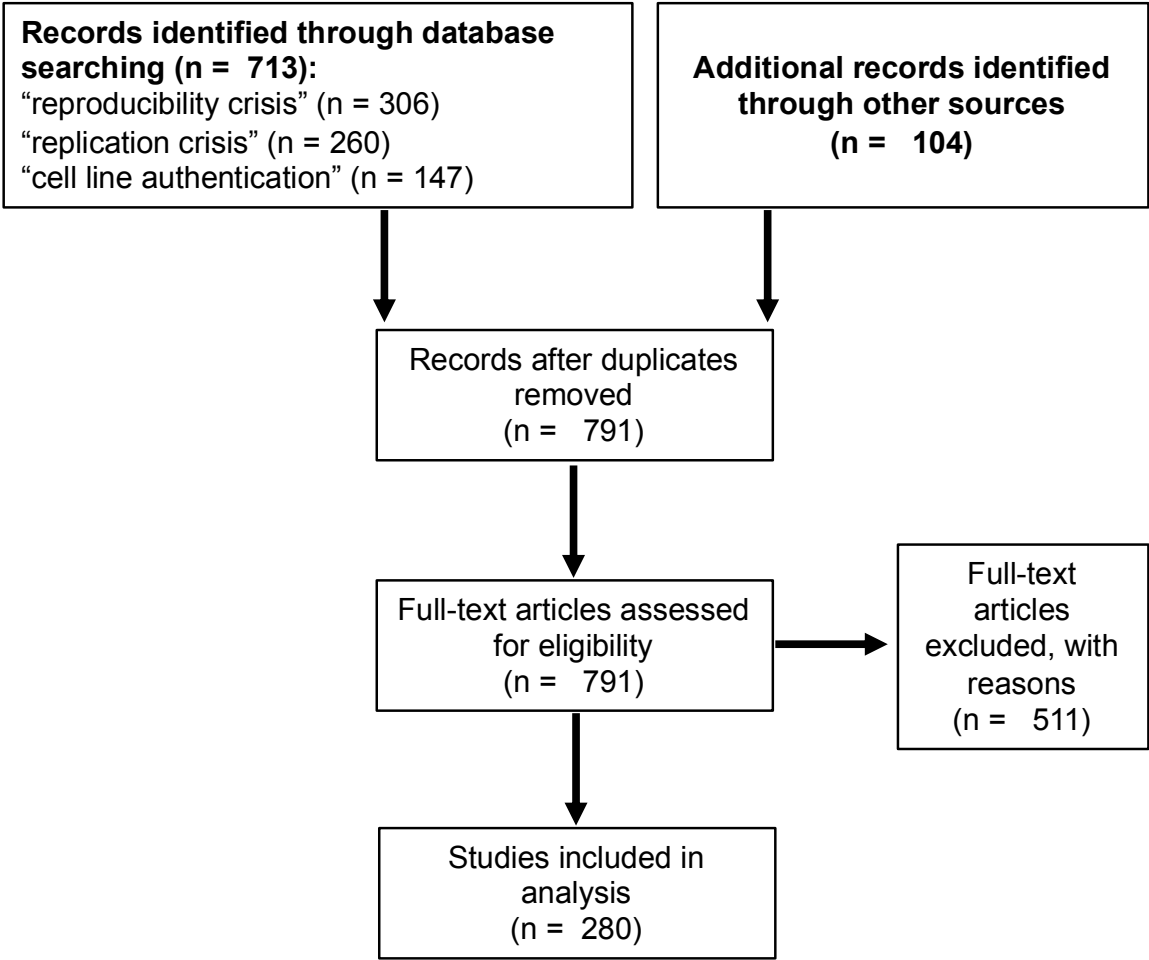


Figure 2. Number of articles that are relevant to the reproducibility crisis and data reproducibility by discipline, article type, and topic. Blue bars indicate the total number of articles. Red bars indicate the number of articles that contain original data.

498

Suppl. Figure 1



Suppl. Figure 1. Summary of the literature search using the search terms “reproducibility crisis”, “replication crisis” (data accessed on 3rd November 2017) PubMed (www.ncbi.nlm.nih.gov/pubmed) to identify articles that report on the reproducibility, replication crisis (data accessed on 3rd November 2017).

499