

Published in final edited form as:

Addiction. 2015 January ; 110(1): 4–8. doi:10.1111/add.12673.

Significance chasing in research practice: Causes, consequences, and possible solutions

Jennifer J. Ware^{1,3,4} and Marcus R. Munafò^{2,3,4}

¹School of Social and Community Medicine, University of Bristol, Bristol, UK

²School of Experimental Psychology, University of Bristol, Bristol, UK

³MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, UK

⁴UK Centre for Tobacco and Alcohol Studies, University of Bristol, Bristol, UK

Abstract

Background and Aims—The low reproducibility of findings within the scientific literature is a growing concern. This may be due to many findings being false positives, which in turn can misdirect research effort and waste money.

Methods—We review factors that may contribute to poor study reproducibility and an excess of ‘significant’ findings within the published literature. Specifically, we consider the influence of current incentive structures, and the impact of these on research practices.

Results—The prevalence of false positives within the literature may be attributable to a number of questionable research practices, ranging from the relatively innocent and minor (e.g., unplanned *post hoc* tests), to the calculated and serious (e.g., fabrication of data). These practices may be driven by current incentive structures (e.g. pressure to publish), alongside the preferential emphasis placed by journals on novelty over veracity. There are a number of potential solutions to poor reproducibility, such as new publishing formats that emphasise the research question and study design, rather than the results obtained. This has the potential to minimise significance chasing and non-publication of null findings.

Conclusions—Significance chasing, questionable research practices, and poor study reproducibility are the unfortunate consequence of a “publish or perish” culture and a preference among journals for novel findings. It is likely that top-down change implemented by those with the ability to modify current incentive structure (e.g., funders and journals) will be required to address problems of poor reproducibility.

While scientists aim to be objective seekers of underlying truths of nature they are also human, and therefore prone to various external influences, personal biases and preconceptions. For example, in order to forge a successful (or indeed any) career as a scientist, one has to publish, preferably regularly and in high Impact Factor (IF) journals. What are the (unintended) consequences of these external influences? There is growing concern that many published scientific findings, across a range of fields, are difficult to

reproduce (1), and may be false (2). This raises the interesting question of whether science is in fact self-correcting, as is typically assumed (3) – once established, false positive findings can become surprisingly difficult to refute (4) – they may become “more ‘vampirical’ than ‘empirical’ – unable to be killed by mere evidence” (5). At the same time, studies that generate null results are often never submitted for publication (6).

How might this situation arise? It is certainly not a new phenomenon – Mendel, for example, famously appears to have dropped observations from his data so that his results conformed to his expectations (7), while Charles Babbage published *Reflections on the Decline of Science in England, and on Some of its Causes* in 1830. Nevertheless, the tools at the disposal of scientists, and the incentive structures within which scientists work, have arguably changed dramatically over recent years. Statistical software packages enable a multiplicity of statistical tests to be conducted, while some countries offer direct financial rewards to scientists who publish in prestigious (i.e., high IF) journals (8), journals which in turn favour studies reporting positive, novel effects over those showing null effects (9). Focusing on statistical significance (as opposed to effect sizes and confidence intervals) exacerbates this problem.

The Science of Reproducibility

A growing “meta-science” literature is beginning to identify factors that contribute to the problems of poor reproducibility. These include study design characteristics which may introduce bias, low statistical power, and flexibility in data collection, analysis, and reporting — termed “researcher degrees of freedom” by Simmons and colleagues (10). Unfortunately detecting the influence of these factors can be difficult because of variability in the reporting of methods and results (11). Nevertheless, some factors are clearly emerging.

One such factor is (low) statistical power. Button and colleagues recently showed that the average power of neuroscience studies is likely to be around 20% (12). Once again, this is not a new problem — Cohen’s classic study on statistical power showed that studies in the 1960 volume of the *Journal of Abnormal and Social Psychology* lacked sufficient power to detect anything other than the largest effects (13). By the time the 1984 volume was published the situation had, if anything, worsened (14). This appears to be due, in part, to lack of appreciation of the importance of statistical power within a Null Hypothesis Significance Testing framework. Vankov and colleagues surveyed the authors of studies published in a high-ranking psychology journal, and found that approximately one third held beliefs that would serve, on average, to reduce statistical power (15). This is critical, since low statistical power increases the likelihood that a statistically significant finding is a false positive (16). Another factor appears to be the country of origin of a study. Studies published in some countries may over-estimate true effects more than those published in other countries (17, 18). This may be because, in certain countries, publication in even medium-rank journals confers substantial direct financial rewards on the authors (19), which may in turn be related to over-estimates of true effects (20).

In most cases, the practices outlined above are not likely to be *conscious* attempts to deceive. However, incidents of purposeful deception have been observed; Diederik Stapel, a

prominent psychologist in the Netherlands with an extensive publication record, is now recognised to have falsified data and wilfully deceived throughout his career (21, 22), while Yoshitaka Fujii, an anaesthesiologist, has had over 100 articles retracted (23). The increase in number of article retractions observed over the past decade (24) suggests that this problem may have become worse in recent years. This increase is seen for both total retractions and retractions due specifically to fraud (i.e., data manipulation or fabrication) (24).

Moreover, journals appear to differ in retraction frequency as a function of IF; Fang & Casadevall (25) observed a robust, positive correlation between journal IF and a “retraction index” (i.e., a measure of retraction frequency). In other words, articles published in high IF journals (those we are supposed to aspire to publish in, and which journal editors hope their journals become) are more likely to be retracted than those published in lower prestige journals. Converging evidence for this effect was reported by Munafò and colleagues (26), who found that journal IF correlates with the extent to which genetic association studies over-estimate the likely underlying true effect. This is what we would expect if publication in these journals, which confers prestige and likely professional success, was influencing the behaviour of scientists and, in extreme cases, fraud. It is possible that articles published in high IF journals are subject to greater scrutiny, which could contribute to higher retraction rates in these journals (25), but Steen and colleagues (27) recently demonstrated that greater scrutiny of high IF publications has had only a “modest” impact on retractions.

While incidents of unambiguous academic fraud fortunately appear to be rare, a more pressing concern is the prevalence of research practices, often well-intentioned or committed unconsciously or unknowingly, which would serve to increase the likelihood that a result is a false positive. A recent study indicated that failing to report all dependent measures employed (63% self-admission rate), selectively reporting only those studies that “worked” (46%), and collecting more data after determining whether the initial results were significant (56%), for example, are all common practices (28). The vast majority of published studies “worked” (i.e., achieved nominal statistical significance), particularly in disciplines such as psychology and psychiatry (29), suggesting that scientific hypothesizing is much more accurate than other forms of precognition (30). The increasing availability of large pre-existing datasets, and the ease with which these can be interrogated for associations likely contributes to this phenomenon.

What are the underlying causes of these problems? Current incentive structures in science (e.g., to publish, particularly in prestigious journals) are perhaps largely responsible. Publication performance is typically linked to institutional funding, and also frequently to rewards at the individual level, such as career progression and salary increases. In some countries, financial rewards are offered to individual scientists who publish in high-profile, international journals (31), which can be in excess of 50% of annual salary (31). This has given rise to concerns that financial incentives may impact negatively on scientific rigour and integrity (8). Indeed, the current peer review process may contribute to the problem, by placing too great an emphasis on novelty over veracity (32).

Increasing Reproducibility

Systemic problems related to incentive structures are, by their nature, difficult to change. Moving from one metric of quality (e.g., IF) to another (e.g., citations) will simply eventually create a different incentive structure to which scientists will (again, consciously or unconsciously) respond to. However, different fields provide examples of potential partial solutions. For example, in clinical trials, pre-registration of study protocols minimises the potential for significance chasing and suppression of null findings. This was introduced in response to evidence that the pharmaceutical industry was less likely to publish null results than favourable results; Etter and colleagues showed publication bias for industry trials of nicotine replacement therapy for smoking cessation, but not for non-industry trials (33). In genetic epidemiology, very large samples, frequently drawn from multiple cohorts, with independent replication and strict statistical criteria for declaring significance are increasingly a minimum requirement for publication (34, 35). In neuroimaging, there is now widespread use of statistical thresholds corrected for multiple comparisons, and/or region of interest analyses specified *a priori*. There is therefore considerable scope for different fields to learn from each other.

It is therefore notable that the recent revision of the Declaration of Helsinki included a small but potentially important change. The pre-registration of research protocols prior to the commencement of data collection is now included as a requirement for *all* research, not just clinical trials as previously (36). This puts journals such as *Addiction* in an interesting position – strictly speaking, authors should have pre-registered their protocols if they are to be compliant with the Declaration of Helsinki, as many journals currently require. In practice, this is likely to be impractical. In particular, it is inappropriate for secondary analyses of existing data sets (where it is difficult or impossible to know whether preliminary analyses were conducted before the protocol was registered). Despite this, many of the concerns around pre-registration are perhaps over-stated. For example, rather than discouraging exploratory research, it simply enables a clearer distinction to be made between exploratory and confirmatory research. There are also no particular reasons why observational studies should not be registered as well as experimental studies (37). Journals such as *Cortex* (38) and *Drug and Alcohol Dependence* (39) have introduced new manuscript submission formats that place the emphasis on the research question and study design, rather than the results obtained. Manuscripts (essentially protocols, containing the introduction, hypotheses, methods, analysis plan and sample size justification) are reviewed *before* data collection takes place, and judged on whether the results will be informative regardless of how they ultimately turn out. If acceptance-in-principle is offered, then the authors can conduct their study safe in the knowledge that, as long as they adhere to their plans, their results will eventually be published.

Deciding on whether or not to publish the results of a study before the results are known offers several important advantages. First, it ensures that publication depends on the importance of the research question being addressed, and the appropriateness of the methods chosen, rather than novelty and *p*-values. Second, it minimises research practices that inflate the likelihood of false positives (e.g., “significance chasing”), given the requirement to adhere to pre-declared methods. Third, the requirement for *a priori* power calculation to

justify the sample size minimises problems of low statistical power (16). Whether this new publication model succeeds remains to be seen, but in principle it could prove a welcome step forward in scientific publishing. *Addiction*, may wish to adopt similar article formats. At the same time, other journals are also adopting new measures to promote transparency of reporting and reproducibility of results. *Nature*, for example, has recently announced the introduction of editorial measures to address concerns regarding reproducibility (40), including a methods reporting checklist (see go.nature.com/oloeip), and the abolition of space restrictions in methods sections.

Conclusions

The prevalence of false positive results in the scientific literature is a growing concern. This may be attributable to a number of practices, ranging from the relatively innocent and minor (e.g., unplanned *post hoc* tests), to the calculated and serious (e.g., fabrication of data). Whatever the intention, the results are at best unhelpful, at worst damaging, and certainly wasteful in terms of time, effort and expense. Whilst such practices are driven by current incentives to publish, alongside the preferential emphasis placed by journals on novelty over veracity (which would continue to drive publication bias even in the absence of poor research practices), positive steps to minimise the risk of false-positives in this field of research can be taken. However, if we believe poor reproducibility is a widespread problem in science, as seems to be the case, we need to consider whether changes to current incentive structures need to be implemented by those with the ability to do so (e.g., funders and journals).

Acknowledgements

The authors wish to thank Chris Chambers for kindly reviewing this manuscript in part prior to submission.

Declarations of Interest: Jennifer Ware is supported by a postdoctoral Research Fellowship from the Oak Foundation. Jennifer Ware and Marcus Munafò are members of the UK Centre for Tobacco and Alcohol Studies, a UK Clinical Research Council Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

1. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014 Jan 30; 505(7485):612–3. [PubMed: 24482835]
2. Ioannidis JP. Why most published research findings are false. *PLoS medicine*. 2005 Aug.2(8):e124. [PubMed: 16060722]
3. Ioannidis J. Why science is not necessarily self-correcting. *Perspect Psychol Sci*. 2012; 7(6):645–54. [PubMed: 26168125]
4. Tatsioni A, Bonitsis NG, Ioannidis JPA. Persistence of contradicted claims in the literature. *Journal of the American Medical Association*. 2007 Dec 5; 298(21):2517–26. [PubMed: 18056905]
5. Freese, J. The problem of predictive promiscuity in deductive applications of evolutionary reasoning to intergenerational transfers: Three cautionary tales. *Intergenerational Caregiving*. Alan Booth, ACC.; Bianchi, Suzanne M.; Seltzer, Judith A., editors. Washington DC: Urban Institute Press; 2008. p. 145-77.

6. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias G. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS one*. 2013; 8(7):e66844. [PubMed: 23861749]
7. Edwards AWF. More on the too-good-to-be-true paradox and Gregor Mendel. *J Hered*. 1986 Mar-Apr;77(2):138. -
8. Fuyuno I, Cyranoski D. Cash for papers: putting a premium on publication. *Nature*. 2006 Jun 15;441(7095):792. [PubMed: 16778850]
9. Brembs B, Button K, Munafo M. Deep impact: unintended consequences of journal rank. *Frontiers in human neuroscience*. 2013; 7:291. [PubMed: 23805088]
10. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*. 2011 Nov; 22(11):1359–66. [PubMed: 22006061]
11. Maynard OM, Munafo MR. Methods reporting in human laboratory studies. *Addiction*. 2013 May; 108(5):1002–3. [PubMed: 23587086]
12. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013 May; 14(5):365–76. [PubMed: 23571845]
13. Cohen J. Statistical power of abnormal-social psychological-research - a review. *J Abnorm Psychol*. 1962; 65(3):145. -&.
14. Sedlmeier P, Gigerenzer G. Do studies of statistical power have an effect on the power of studies. *Psychol Bull*. 1989 Mar; 105(2):309–16.
15. Vankov I, Bowers J, Munafo MR. On the persistence of low power in psychological science. *Quarterly journal of experimental psychology*. 2014 Mar 3;
16. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews Neuroscience*. 2013 May; 14(5):365–76.
17. Fanelli D, Ioannidis JPA. US studies may overestimate effect sizes in softer research. *Proceedings of the National Academy of Sciences USA*. 2013
18. Munafo MR, Attwood AS, Flint J. Bias in genetic association studies: effects of research location and resources. *Psychol Med*. 2008 Aug; 38(8):1213–4. [PubMed: 18598581]
19. Shao JF, Shen HY. The outflow of academic papers from China: why is it happening and can it be stemmed? *Learn Publ*. 2011 Apr; 24(2):95–7.
20. Pan ZL, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JPA. Local literature bias in genetic epidemiology: An empirical evaluation of the Chinese literature. *Plos Med*. 2005 Dec; 2(12):1309–17.
21. Carey B. Fraud case seen as a red flag for psychology research. *New York Times*. 2011
22. Stroebe W, Postmes T, Spears R. Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*. 2012; 7(6):670–88. [PubMed: 26168129]
23. Cyranoski D. Retraction record rocks community. *Nature*. 2012 Sep 20; 489(7416):346–7. [PubMed: 22996524]
24. Steen RG. Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of Medical Ethics*. 2011 Apr; 37(4):249–53. Epub 2010/12/28. eng. [PubMed: 21186208]
25. Fang FC, Casadevall A. Retracted science and the retraction index. *Infection and Immunity*. 2011 Oct; 79(10):3855–9. Epub 2011/08/10. eng. [PubMed: 21825063]
26. Munafo MR, Stohart G, Flint J. Bias in genetic association studies and impact factor. *Molecular Psychiatry*. 2009 Feb; 14(2):119–20. Epub 2009/01/22. eng. [PubMed: 19156153]
27. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS one*. 2013; 8(7):e68397. [PubMed: 23861902]
28. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*. 2012 May 1; 23(5):524–32. [PubMed: 22508865]
29. Fanelli D. "Positive" results increase down the Hierarchy of the Sciences. *PLoS one*. 2010; 5(4):e10068. [PubMed: 20383332]

30. Bones AK. We knew the future all along: Scientific hypothesizing is much more accurate than other forms of precognition - a satire in one part. *Perspect Psychol Sci.* 2012; 7(3):307–9. [PubMed: 26168468]
31. Franzoni C, Scellato G, Stephan P. Science policy. Changing incentives to publish. *Science.* 2011 Aug 5; 333(6043):702–3. [PubMed: 21817035]
32. Yong E. Replication studies: Bad copy. *Nature.* 2012 May 17; 485(7398):298–300. [PubMed: 22596136]
33. Etter JF, Burri M, Stapleton J. The impact of pharmaceutical company funding on results of randomized trials of nicotine replacement therapy for smoking cessation: a meta-analysis. *Addiction.* 2007 May; 102(5):815–22. [PubMed: 17493109]
34. Munafò MR, Gage SH. Improving the reliability and reporting of genetic association studies. *Drug Alcohol Depend.* 2013 Apr 26. Epub 2013/05/01. Eng.
35. Munafò MR. Reliability and replicability of genetic association studies. *Addiction.* 2009 Sep; 104(9):1439–40. [PubMed: 19686515]
36. World-Medical-Association: World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. *JAMA : the journal of the American Medical Association.* 2013 Oct 19.
37. Dal-Re R, Ioannidis JP, Bracken MB, Buffler PA, Chan AW, Franco EL, et al. Making prospective registration of observational research a reality. *Science translational medicine.* 2014 Feb 19.6(224):224cm1.
38. Chambers CD. Registered reports: a new publishing initiative at Cortex. *Cortex; a journal devoted to the study of the nervous system and behavior.* 2013 Mar; 49(3):609–10.
39. Munafò MR, Strain E. Registered Reports: A new submission format at Drug and Alcohol Dependence. *Drug and alcohol dependence.* 2014
40. Nature. Announcement: Reducing our irreproducibility. *Nature.* 2013; 496:398.