# Deep Impact: Unintended consequences of journal rank

Björn Brembs[1] and Marcus Munafò[2]

1. Institute of Zoology – Neurogenetics, University of Regensburg, Universitätsstr. 31, 93040 Regensburg, Germany, bjoern@brembs.net

2. School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom.

http://arxiv.org/abs/1301.3748

Corresponding Author: Björn Brembs

Abstract

Much has been said about the increasing bureaucracy in science, stifling innovation, hampering the creativity of researchers and incentivizing misconduct, even outright fraud. Many anecdotes have been recounted, observations described and conclusions drawn about the negative impact of impact assessment on scientists and science. However, few of these accounts have drawn their conclusions from data, and those that have typically relied on a few studies. In this review, we present the most recent and pertinent data on the consequences that our current scholarly communication system has had on various measures of scientific quality (such as utility/citations, methodological soundness, expert ratings and retractions). These data confirm previous suspicions: using journal rank as an assessment tool is bad scientific practice. Moreover, the data lead us to argue that any journal rank (not only the currently-favored Impact Factor) would have this negative impact. Therefore, we suggest that abandoning journals altogether, in favor of a library-based scholarly communication system, will ultimately be necessary. This new system will use modern information technology to vastly improve the filter, sort and discovery function of the current journal system.

# Introduction

Science is the bedrock of modern society, improving our lives through advances in medicine, communication, transportation, forensics, entertainment and countless other areas. Moreover, today's global problems cannot be solved without scientific input and understanding. The more our society relies on science, and the more our population becomes scientifically literate, the more important the reliability (i.e., veracity and integrity, or, 'credibility' [1]) of scientific research becomes. Scientific research is largely a public endeavor, requiring public trust. Therefore, it is critical that public trust in science remains high. In other words, the reliability of science is not only a societal imperative, it is also vital to the scientific community itself. However, every scientific publication may in principle report results which prove to be unreliable, either unintentionally, in the case of honest error or statistical variability, or intentionally in the case of misconduct or fraud. Even under ideal circumstances, science can never provide us with absolute truth. In Karl Popper's words: "Science is not a system of certain, or established, statements" [2]. Peer-review is one of the mechanisms which have evolved to increase the reliability of the scientific literature.

At the same time, the current publication system is being used to structure the careers of the members of the scientific community by evaluating their success in obtaining publications in high-ranking journals. The hierarchical publication system ('journal rank') used to communicate scientific results is thus central not only to the composition of the scientific community at large (by selecting its members), but also to science's position in society. In recent years, the scientific study of the effectiveness of such measures of quality control has grown.

# Retractions and the Decline Effect

A disturbing trend has recently gained wide public attention: The retraction rate of articles published in scientific journals, which had remained stable since the 1970's, began to increase rapidly in the early 2000's from 0.001% of the total to about 0.02% (Figure 1a). In 2010 we have seen the creation and popularization of a website dedicated to monitoring retractions (http://retractionwatch.com), while 2011 has been described as the "the year of the retraction" [3]. The reasons suggested for retractions vary widely, with the recent sharp rise potentially facilitated by an increased willingness of journals to issue retractions, or increased scrutiny and error-detection from online media. Although cases of clear scientific misconduct initially constituted a minority of cases [4–9], the fraction of retractions due to misconduct has risen sharper than the overall retraction rate and now the majority of all retractions is due to misconduct [10,11].

Retraction notices, a metric which is relatively easy to collect, only constitute the extreme end of a spectrum of unreliability that is inherent to the scientific method: we can never be *entirely* certain of our results [2]. Much of the training scientists receive aims to reduce this uncertainty long before the work is submitted for publication. However, a less readily quantified but more frequent phenomenon (compared to rare retractions) has recently garnered attention, which calls into question the effectiveness of this training. The 'decline-effect', which is now well-described, relates to the observation that the strength of evidence for a particular finding often declines over time [12–22]. This effect provides wider scope for assessing the unreliability of scientific research than retractions alone, and allows for more general conclusions to be drawn.

Researchers make choices about data collection and analysis which increase the chance of false-positives (i.e., researcher bias) [18,19], and surprising and novel effects are more likely to be published than studies showing no effect. This is the well-known phenomenon of publication bias [12,20–26]. In other words, the probability of getting a paper published might be biased towards larger initial effect sizes, which are revealed by later studies to be not so large (or even absent entirely), leading to the so-called decline effect. While sound methodology can help reduce researcher bias [18], publication bias is more difficult to address. Some journals are devoted to publishing null results, or have sections devoted to these, but none is particularly high-ranking or well-read [12,27]. Publication therein is typically not a cause for excitement [27,28], leading to an overall low frequency of replications [29–33]. Publication bias is also exacerbated by a tendency for journals to be less likely to publish replication studies (or, worse still, failures to replicate) [29,32,34–37]. Here we argue that the counter-measures proposed to improve the reliability and veracity of science such as peer-review in a hierarchy of journals or methodological training of scientists may not be sufficient – and hence laud recent efforts, such as the 'Reproducibility Initiative' [38] or the "Reproducibility Project" [39].

While there is growing concern regarding the increasing rate of retractions in particular, and the unreliability of scientific findings in general, little consideration has been given to the infrastructure by which scientists not only communicate their findings but also evaluate each other as a potential contributing factor. That is, to what extent does the environment in which science takes place contribute to the problems described above? By far the most common metric by which publications are evaluated, at least initially, is the perceived prestige or rank of the journal in which they appear. Does the pressure to publish in prestigious, high-ranking journals contribute to the unreliability of science?

# The Decline Effect and Journal Rank

The common pattern seen where the decline effect has been documented is one of an initial publication in a high-ranking journal, followed by attempts at replication in lower-ranked journals which either failed to replicate the original findings, or suggested a much weaker effect [13]. Journal rank is most commonly assessed using Thomson Reuters' Impact Factor (IF), which has been shown to correspond well with subjective ratings of journal quality and rank [40–43]. One particular case [24] illustrates the decline effect (Figure 1b), and shows that early publications both report a larger effect than subsequent studies, and are also published in journals with a higher IF. These observations raise the more general question of whether research published in high-ranking journals is inherently less reliable than research in lower-ranking journals.

As journal rank is also predictive of the incidence of fraud and misconduct in retracted publications, as opposed to other reasons for retraction [44], it is not surprising that higher ranking journals are also more likely to publish fraudulent work than lower ranking journals [10]. These data, however, cover only the small fraction of publications that have been retracted. More important is the large body of the literature that is not retracted and thus actively being used by the scientific community. There is evidence that unreliability is higher in high-ranking journals as well, also for non-retracted publications: A meta-analysis of genetic association studies provides evidence that the extent to which a study over-estimates the likely true effect size is positively correlated with the IF of the journal in which it is published (Figure 1c) [45]. Similar effects have been reported in the context of other research fields [46–48]. It is therefore not surprising that journal rank is an strong predictor of the rate of retractions (Figure 1d) [8,49,50].

There are several converging lines of evidence which indicate that publications in high ranking journals are not only more likely to be fraudulent than articles in lower ranking journals, but also more likely to present discoveries which are less reliable (i.e., are inflated, or cannot subsequently be replicated). Some of the sociological mechanisms behind these correlations have been documented, such as pressure to publish (preferably positive results in high-ranking journals), leading to the potential for decreased ethical standards [51] and increased publication bias in highly competitive fields [16]. The general increase in competitiveness, and the precariousness of scientific careers [52], may also lead to an increased publication bias across the sciences [53]. This evidence supports earlier propositions about social pressure being a major factor driving misconduct and publication bias [54], eventually culminating in retractions in the most extreme cases.

Some of the relationship between number of retractions and journal rank may be explained by the greater visibility of publications in these journals, which is both one of the incentives driving publication bias, and a likely underlying cause for the detection of error or misconduct with the eventual retraction of the publications as a result [8]. Conversely, the scientific community may also be less concerned about incorrect findings published in more obscure journals. The finding that most retractions come from the numerous lower-ranking journals [10] casts doubt on such reasoning. Moreover, other evidence also suggests comparatively low visibility effects: If the greater visibility of publications in high-ranking journals were responsible for many of the retractions, one would expect at least an equally strong, if not stronger effect on other potential measures of quality such as citations, expert opinion, reproducibility or methodological standards. However, many of these measures fail to show any correlation with journal rank and those effects that are observed are much weaker than one would expect, given the data described above.

# Journal Rank and Study Impact

Thus far we have presented evidence that research published in high-ranking journals may be less reliable and thus at a higher risk of being retracted, compared with publications in lower-ranking journals. Nevertheless, there is a strong common perception that high-ranking journals publish ‘better’ or ‘more important’ science, and that the IF captures this well [40,41]. The assumption is that high-ranking journals are able to be highly selective and publish only the most important and best-supported scientific discoveries, which will then, as a consequence of their quality, go on to be highly cited [25]. One way to reconcile this common perception with the data would be that while journal rank may be indicative of a minority of unreliable publications, it may also (or more strongly) be indicative of scientific quality in the majority of remaining, reliable publications. Given this, one would expect three things: 1) publications in high-ranking journals should be cited more often than publications in low-ranking journals; 2) expert ratings of published articles should correlate well with journal rank; and, 3) other potential measures of scientific quality, such as adherence to methodological standards or reproducibility, should also correlate well with journal rank. All three of these predictions are challenged by the available data.

Beyond the quality of the research, there are three additional reasons why publications in high-ranking journals might receive a high number of citations. First, publications in high-ranking journals achieve greater exposure by virtue of the larger circulation of the journal in which they appear. Second, citing high-ranking publications in one's own publication may increase its perceived value. Third, the surprising, counter-intuitive or controversial findings often published in high-ranking journals, draw citations not only from follow-up studies but also from news-type articles discussing the controversy. Despite these four factors, it has been established for some time that journal rank is a measurable, but unexpectedly weak predictor of future citations [26,55–59]. The data presented in a recent analysis of the development of these correlations between journal rank and future citations over the period from 1902-2009 reveal two very informative trends [60]. First, while the predictive power of journal rank remained very low for the entire first two thirds of the 20$^{th}$ century, it started to slowly increase shortly after the publication of the first IF data in the 1960's. This correlation kept increasing until the second interesting trend emerged with the advent of the internet and keyword-search in the 1990's, from which time on it fell back to pre-1960's levels until the end of the study period in 2009. Overall, consistent with the citation data already available, the coefficient of determination between journal rank and citations was always in the range of

~0.1 to 0.3 (i.e., very low). It thus appears that a large part of what little correlation between journal rank and citations can be observed, indeed stems from visibility effects due to the influence of the IF on reading habits rather than from factors intrinsic to the published articles. Supporting these weak correlations are data reporting classification errors (i.e., whether a publication received too many or too few citations with regard to the rank of the journal it was published in) at or exceeding 30% [57,58,61,62].

The only measure of citation count that does correlate strongly with journal rank (negatively) is the number of articles without any citations at all [63], supporting the argument that fewer articles in high-ranking journals go unread. Thus, while the data corroborate the hypothesis that visibility does to some extent affect citations (and by extension presumably also retractions [8]), it is difficult to explain why this effect should be larger for retractions than for citations, rather than the reverse. Even the assumption that selectivity might confer a citation advantage is challenged by evidence that, in the citation analysis by Google Scholar, only the most highly selective journals such as *Nature* and *Science* come out ahead over unselective preprint repositories such as ArXiv and RePEc (Research Papers in Economics) [64].

Expert ratings are another means by which a study's impact can be assessed, and the correlation between such ratings and journal rank is strikingly similar to that observed for citations (that is, observable but much lower than one would expect if the high correlation with retractions were due to a large visibility effect) [65]. Other measures might gauge scientific quality more directly. Adherence to basic principles of sound scientific methodology (c.f., the CONSORT statement: http://www.consort-statement.org), or the extent to which the published results can be replicated in other laboratories, would qualify as such metrics. Three different studies on levels of evidence in medical research have found varying results. While two studies on surgery journals found a correlation between IF and the levels of evidence defined in the respective studies [66,67], a study of anesthesia journals failed to find any statistically significant correlation between journal rank and evidence-based medicine principles [68]. The two surgery studies covered an IF range between 0.5 and 2.0, and 0.7 and 1.2, while the anesthesia study covered the range 0.8 to 3.5, so that it is possible that any correlation at the lower end of the scale is abolished when higher rank journals are included. Clearly, more evidence is required to establish whether journal rank is a predictor of methodological soundness.

Beyond expert ratings and sound methodology, reproducibility is at the core of the scientific method and thus a hallmark of scientific quality. Three recent studies reported attempts to replicate published findings in preclinical medicine [69–71]. All three found a very low frequency of reproduction, suggesting that maybe only one out of five preclinical findings is reproducible. In fact, the level of reproducibility was so low that no relationship between journal rank and reproducibility could be detected. With several independent measures failing to provide compelling evidence that journal rank is a reliable predictor of scientific impact or quality, and other measures indicating that journal rank is at least equally if not more predictive of low reliability, the central role of journal rank in modern science deserves close scrutiny.

# Practical consequences of Journal Rank

Even if a particular study has been performed to the highest standards, the quest for publication in high-ranking journals slows down the dissemination of science and increases the burden on reviewers, by iterations of submissions and rejections cascading down the hierarchy of journal rank [57,72,73]. A recent study seems to suggest that such rejections eventually improve manuscripts enough to yield measurable citation benefits [74]. However, the effect size of such resubmissions appears to be on the order of 0.1 citations per article, a statistically significant, but in practical terms negligible effect. This conclusion is corroborated by an earlier study which did not find such an effect [73].

More perniciously, journal rank may distort the record of discoveries and their attribution to individual scientists. For instance, the recent discovery of a 'Default-Mode Network' in rodent brains was, presumably, made independently by two different sets of neuroscientists and published only within a few months of each other [75,76]. Perhaps because of journal rank, the later publication in the higher ranking journal [75] was mentioned in a subsequent high-ranking publication [77]. It is straightforward to project that the later publication will now go on to be cited more often than the earlier report of the same discovery in a lower ranking journal [76], especially since the later publication did not cite the earlier one, despite the final version having been submitted months after the earlier publication appeared. We do not know of any empirical studies quantitatively addressing this particular effect of journal rank.

Finally, the focus on journal rank has also allowed corporate publishers to keep their most prestigious journals closed-access and to increase subscription prices [78], creating additional barriers to the dissemination of science. The argument from highly selective journals is that their per-article cost would be too high for author processing fees, which may be up to 37,000€ (US$48,000) for the journal *Nature* [79].

# Impact Factor – Negotiated, irreproducible and unsound

The IF is a metric for the number of citations to articles in a journal (the numerator), normalized by the number of articles in that journal (the denominator). However, there is evidence that IF is, at least in some cases, not calculated but negotiated, that it is not reproducible, and that, even if it were reproducibly computed, the way it is derived is not mathematically sound. The fact that publishers have the option to negotiate how their IF is calculated is well-established – in the case of *PLoS Medicine*, the negotiation range was between 2 and about 11 [80]. What is negotiated is the denominator in the IF equation (i.e., which published articles which are counted), given that all citations count towards the numerator whether they result from publications included in the denominator or not. Removing editorials and News-and-Views articles from the denominator (so called “front-matter”) can therefore dramatically alter the resulting IF [81–85]. While these IF negotiations between are rarely made public, the number of citations (numerator) and published articles (denominator) used to calculate IF are accessible via *Journal Citation Reports*. This database can be searched for evidence that the IF has been negotiated. For instance, the numerator and denominator values for *Current Biology* in 2002 and 2003 indicate that while the number of citations remained relatively constant, the number of published articles dropped. This decrease occurred after the journal was purchased by Cell Press (an imprint of Elsevier), despite there being no change in the layout of the journal. Critically, the arrival of a new publisher corresponded with a *retrospective* change in the denominator used to calculate IF (Table 1).

In an attempt to test the accuracy of the ranking of some of their journals by IF, Rockefeller University Press purchased access to the citation data of their journals and some competitors. They found numerous discrepancies between the data they received and the published rankings, sometimes leading to differences of up to 19% [86]. When asked to explain this discrepancy, Thomson Reuters replied that they routinely use several different databases and had accidentally sent Rockefeller University Press the wrong one. Despite this, a second database sent also did not match the published records. This is only one of a number reported errors and inconsistencies [87,88].

It is well-known that citation data are strongly left-skewed, meaning that a small number of publications receive a large number of citations, while most publications receive very few [56–58,63,81,86,89,90]. The use of an arithmetic mean as a measure of central tendency on such data (rather than, say, the median) is clearly inappropriate, but this is exactly what is used in the IF calculation. The International Mathematical Union reached the same conclusion in an analysis of the IF [91].

Complementing the specific flaws just mentioned, a recent, comprehensive review of the bibliometric literature lists various additional shortcomings of the IF more generally [92].

# Conclusions

While at this point it seems impossible to quantify the relative contributions of the different factors influencing the reliability of scientific publications, the current empirical literature on the effects of journal rank provides evidence supporting the following four conclusions: 1) Journal rank is a weak to moderate predictor of scientific impact; 2) Journal rank is a moderate to strong predictor of both intentional and unintentional scientific unreliability; 3) Journal rank is expensive, delays science and frustrates researchers; and, 4) Journal rank as established by IF violates even the most basic scientific standards, but predicts subjective judgments of journal quality.

# Caveats

While the latter two conclusions appear uncontroversial, the former two are counter-intuitive and require explanation. Weak correlations between citations and journal rank based on IF may be caused by the poor statistical properties of the IF. This explanation could (and should) be tested by using any of the existing alternative ranking tools available (such as Thomson Reuters' Eigenfactor or Scopus' SCImagoJournalRank, etc.) and computing correlations with the metrics discussed above. Alternatively, one can choose important metrics and compute which journals score particularly high on these. However, since the IF reflects the common perception of journal hierarchies rather well [40–43], any alternative hierarchy that would better reflect article citation frequencies would likely violate this intuitive sense of journal rank, as different ways to compute journal rank lead to different hierarchies [93]. Both alternatives thus challenge our subjective journal ranking.

This subjective ranking of journals also leads to a circularity that confounds many empirical studies. That is, authors use journal rank, in part, to make decisions of where to submit their manuscripts, such that well-performed studies yielding ground-breaking discoveries with general implications are preferentially submitted to high-ranking journals. Readers, in turn, expect only to read about such articles in high-ranking journals, leading to the exposure and visibility confounds discussed above and at length in the cited literature. Moreover, citation practices and methodological standards vary in different scientific fields, potentially distorting both the citation and reliability data. Given these confounds one might expect highly varying and often inconclusive results. Despite this, the literature contains many studies showing associations between journal rank on several measures of scientific quality, but also contains at least equally strong, consistent effects of journal rank predicting scientific unreliability. Neither group of studies can thus be easily dismissed, suggesting that the incentives journal rank creates for the scientific community (to submit either their best or their most unreliable work to the most high-ranking journals) at best cancel each other out. Such unintended consequences are well-known from other fields where metrics are applied [94].

Thus, while there are concerns not only about the validity of the IF to adequately capture journal rank but also about confounding factors complicating the interpretation of some of the data, we find, in the absence of additional data, that these concerns do not suffice to substantially question our conclusions, but do emphasize the need for future research.

# Potential long-term consequences of journal rank

Taken together, the reviewed literature suggests that using journal rank is unhelpful at best and unscientific at worst. In our view, IF generates an illusion of exclusivity and prestige based on an assumption that it will predict subsequent impact, which is not supported by empirical data. As the IF aligns well with intuitive notions of journal hierarchies [40–42], it receives insufficient scrutiny [95] (perhaps a case of confirmation bias). The one field in which journal rank is scrutinized is bibliometrics. We have reviewed the pertinent empirical literature to supplement the largely argumentative discussion on the opinion pages of many learned journals [12,54,81,85,90,96–105] with empirical data. Journal rank seems to appeal to subjective impressions of certain effects, but these effects disappear as soon as they are subjected to scientific scrutiny.

In our understanding of the data, the social and psychological influences described above are, at least to some extent, generated by journal rank itself, which in turn may contribute to the observed decline effect and rise in retraction rate. That is, systemic pressures on the author, rather than increased scrutiny on the part of the reader, inflate the unreliability of much scientific research. Without reform of our publication system, the incentives associated with increased pressure to publish in high-ranking journals will continue to encourage scientists to be less cautious in their conclusions (or worse), in an attempt to market their research to the top journals [16,45,51,52,54]. This is reflected in the decline in null results reported across disciplines and countries [53], and corroborated by the findings that much of the increase in retractions may be due to misconduct [10,11], and that much of this misconduct occurs in studies published high-ranking journals [10,44]. Inasmuch as journal rank guides the appointment and promotion policies of research institutions, the increasing rate of misconduct that has recently been observed may prove to be but the beginning of a pandemic: It is conceivable that, for the last few decades, research institutions world-wide may have been hiring and promoting scientists who excel at marketing their work to top journals, but who are not necessarily equally good at conducting their research. Conversely, these institutions may have purged excellent scientists from their ranks, whose marketing skills did not meet institutional requirements. If this interpretation of the data is correct, we now have a generation of excellent marketers (possibly, but not necessarily also excellent scientists) as the leading figures of the scientific enterprise, constituting another potentially major contributing factor to the rise in retractions. This generation is now in charge of training the next generation of scientists, with all the foreseeable consequences for the reliability of scientific publications in the future.

The implications of the data presented here go beyond the reliability of scientific publications – public trust in science and scientists has been in decline for some time in many countries [106–108], dramatically so in some sections of society [109], culminating in the sentiment that scientists are nothing more than yet another special interest group [110,111]. In the words of Daniel Sarewitz: “Nothing will corrode public trust more than a creeping awareness that scientists are unable to live up to the standards that they have set for themselves” [105]. The data presented here prompt the suspicion that the corrosion has already begun and that journal rank may have played a part in this decline as well.

# Alternatives

Alternatives to journal rank exist – we now have technology at our disposal which allows us to perform all of the functions journal rank is currently supposed to perform in an unbiased, dynamic way on a per-article basis, allowing the research community greater control over selection, filtering, and ranking of scientific information [57,112–115]. Since there is no technological reason to continue using journal rank, one implication of the data reviewed here is that we can instead use current technology and remove the need for a journal hierarchy completely. As we have argued, it is not only technically obsolete, but also counter-productive and a potential threat to the scientific endeavor. We therefore would favor bringing scholarly communication back to the research institutions in an archival publication system in which both software, raw data and their text descriptions are archived and made accessible, after peer-review and with scientifically-tested metrics accruing reputation in a constantly improving reputation system [116]. This reputation system would be subjected to the same standards of scientific scrutiny as are commonly applied to all scientific matters and evolve to minimize gaming and maximize the alignment of researchers' interests with those of science (which are currently misaligned [27]). Only an elaborate ecosystem of a multitude of metrics can provide the flexibility to capitalize on the small fraction of the multi-faceted scientific output that is actually quantifiable. Funds currently spent on journal subscripts could easily suffice to finance the initial conversion of scholarly communication, even if only as long-term savings. Other solutions certainly exist [73,117], but the need for an alternative system is clearly pressing [118].

Importantly, the three models which are currently aimed at publishing reform are not sustainable in the long term. First, **Gold Open Access** publishing without abolishment of journal rank (or strong market regulation with, e.g. strict price caps) will lead to a luxury segment in the market, as evidenced not only by suggested article processing charges nearing 40,000€ (US$50,000) for the highest-ranking journals [79], but also by the correlation of existing article processing charges with journal rank [119]. Such a luxury segment would entail that only the most affluent institutions or authors would be able to afford publishing their work in high-ranking journals, anathema to the supposed meritocracy of science. Hence, universal, unregulated Gold Open Access is one of the few situations we can imagine that would potentially be even worse than the current *status quo*. Second, **Green Open Access** publishing, while expected to be more cost-effective for institutions than Gold Open Access [120], entails twice the work on the part of the authors and needs to be mandated and enforced to be effective, thus necessitating an additional layer of bureaucracy, on top of the already unsustainable status quo, which would not be seriously challenged. Moreover, some publishers have excluded any cooperation with green publishing schemes. Third, **Hybrid Open Access** publishing inflates pricing [119] and allows publishers to not only double-dip into the public purse, but to triple-dip. Thus, Hybrid Open Access publishing is probably the most expensive option overall.

In conclusion, the *status quo* of scholarly communication is a threat to the scientific endeavor and the three models currently vying to replace it are not sustainable, either. In our opinion, reform is needed.

Acknowledgements

Neil Saunders was of tremendous value in helping us obtain and understand the PubMed retraction data for Figure 1a. We are indebted to John Ioannidis, Daniele Fanelli, Christopher Baker, Dwight Kravitz, Tom Hartley, Jason Priem, Stephen Curry and three anonymous reviewers for their comments on an earlier version of this manuscript. MRM is a member of the UK Centre for Tobacco Control Studies, a UKCRC 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. BB was a Heisenberg-Fellow of the DFG during the time most of this manuscript was written and their support is gratefully acknowledged as well.

# References


1. Ioannidis JPA (2012) Why Science Is Not Necessarily Self-Correcting. Perspectives on Psychological Science 7: 645–654. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691612464056. Accessed 8 November 2012.

2. Popper K (1995) In Search of a Better World: Lectures and Essays from Thirty Years. Routledge; New edition edition (December 20, 1995).

3. Hamilton J (2011) Debunked Science: Studies Take Heat In 2011. NPR. Available: http://www.npr.org/2011/12/29/144431640/debunked-science-studies-take-heat-in-2011. Accessed 8 March 2012.

4. Fanelli D (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PloS one 4: e5738. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2685008&tool=pmcentrez&rendertype=abstract. Accessed 12 March 2012.

5. Van Noorden R (2011) Science publishing: The trouble with retractions. Nature 478: 26–28. Available: http://www.ncbi.nlm.nih.gov/pubmed/21979026. Accessed 14 March 2012.

6. Wager E, Williams P (2011) Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. Journal of medical ethics 37: 567–570. Available: http://www.ncbi.nlm.nih.gov/pubmed/21486985. Accessed 14 March 2012.

7. Nath SB, Marcus SC, Druss BG (2006) Retractions in the research literature: misconduct or mistakes? The Medical journal of Australia 185: 152–154. Available: http://www.ncbi.nlm.nih.gov/pubmed/16893357. Accessed 14 March 2012.

8. Cokol M, Iossifov I, Rodriguez-Esteban R, Rzhetsky A (2007) How many scientific papers should be retracted? EMBO reports 8: 422–423. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1866214&tool=pmcentrez&rendertype=abstract. Accessed 14 March 2012.

9. Steen RG (2011) Retractions in the scientific literature: do authors deliberately commit research fraud? Journal of medical ethics 37: 113–117. Available: http://www.ncbi.nlm.nih.gov/pubmed/21081306. Accessed 30 March 2012.

10. Fang FC, Steen RG, Casadevall A (2012) Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences of the United States of America 109: 17028–17033. Available: http://www.pnas.org/cgi/content/long/109/42/17028. Accessed 1 November 2012.

11. Steen RG (2011) Retractions in the scientific literature: is the incidence of research fraud increasing? Journal of medical ethics 37: 249–253. Available: http://www.ncbi.nlm.nih.gov/pubmed/21186208. Accessed 14 March 2012.

12. Schooler J (2011) Unpublished results hide the decline effect. Nature 470: 437. Available: http://www.ncbi.nlm.nih.gov/pubmed/21350443. Accessed 8 March 2012.

13. Lehrer J (2010) The decline effect and the scientific method. New Yorker. Available: http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer. Accessed 8 March 2012.

14. Bertamini M, Munafo MR (2012) Bite-Size Science and Its Undesired Side Effects. Perspectives on Psychological Science 7: 67–71. Available: http://pps.sagepub.com/content/7/1/67.full. Accessed 8 March 2012.

15. Palmer AR (2000) QUASI-REPLICATION AND THE CONTRACT OF ERROR: Lessons from Sex Ratios, Heritabilities and Fluctuating Asymmetry. Annual Review of Ecology and Systematics 31: 441–480. Available: http://www.annualreviews.org/doi/pdf/10.1146/annurev.ecolsys.31.1.441. Accessed 8 March 2012.

16. Fanelli D (2010) Do pressures to publish increase scientists' bias? An empirical support from US States Data. PloS one 5: e10271. Available: http://dx.plos.org/10.1371/journal.pone.0010271. Accessed 2 March 2012.

17. Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/.

18. Simmons LW, Tomkins JL, Kotiaho JS, Hunt J (1999) Fluctuating paradigm. Proceedings of the Royal Society B: Biological Sciences 266: 593–595. Available: http://rspb.royalsocietypublishing.org/cgi/content/abstract/266/1419/593. Accessed 8 March 2012.

19. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science 22: 1359–1366. Available: http://www.ncbi.nlm.nih.gov/pubmed/22006061. Accessed 8 March 2012.

20. Møller AP, Jennions MD (2001) Testing and adjusting for publication bias. Trends in Ecology & Evolution 16: 580–586. Available: http://dx.doi.org/10.1016/S0169-5347(01)02235-2. Accessed 8 March 2012.

21. Møller AP, Thornhill R, Gangestad SW (2005) Direct and indirect tests for publication bias: asymmetry and sexual selection. Animal Behaviour 70: 497–506. Available: http://dx.doi.org/10.1016/j.anbehav.2004.11.017. Accessed 8 March 2012.

22. Van Dongen S (2011) Associations between asymmetry and human attractiveness: Possible direct effects of asymmetry and signatures of publication bias. Annals of human biology 38: 317–323. Available: http://informahealthcare.com/doi/abs/10.3109/03014460.2010.544676. Accessed 8 March 2012.

23. Song F, Eastwood A, Gilbody S, Duley L (1999) The role of electronic journals in reducing publication bias. Informatics for Health and Social Care 24: 223–229.

Available: http://informahealthcare.com/doi/abs/10.1080/146392399298429. Accessed 22 February 2012.

24. Munafò MR, Matheson IJ, Flint J (2007) Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. Molecular psychiatry 12: 454–461. Available: http://www.ncbi.nlm.nih.gov/pubmed/17453061. Accessed 8 March 2012.

25. Young NS, Ioannidis JPA, Al-Ubaydli O (2008) Why current publication practices may distort science. PLoS medicine 5: e201. Available: http://dx.plos.org/10.1371/journal.pmed.0050201. Accessed 8 March 2012.

26. Callaham M (2002) Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals. JAMA: The Journal of the American Medical Association 287: 2847–2850. Available: http://jama.ama-assn.org/cgi/doi/10.1001/jama.287.21.2847. Accessed 14 March 2012.

27. Nosek BA, Spies JR, Motyl M (2012) Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. Perspectives on Psychological Science 7: 615–631. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691612459058. Accessed 8 November 2012.

28. Giner-Sorolla R (2012) Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science. Perspectives on Psychological Science 7: 562–571. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691612457576. Accessed 8 November 2012.

29. Hartshorne JK, Schachner A (2012) Tracking Replicability as a Method of Post-Publication Open Evaluation. Frontiers in Computational Neuroscience 6: 8. Available: http://www.frontiersin.org/Computational_Neuroscience/10.3389/fncom.2012.00008/abstract. Accessed 5 March 2012.

30. Kelly CD (2006) Replicating Empirical Research In Behavioral Ecology: How And Why It Should Be Done But Rarely Ever Is. The Quarterly Review of Biology 81: 221–236. Available: http://www.jstor.org/stable/10.1086/506236. Accessed 8 March 2012.

31. Carpenter S (2012) Psychology's Bold Initiative. Science 335: 1558–1561. doi:10.1126/science.335.6076.1558.

32. Yong E (2012) Replication studies: Bad copy. Nature 485: 298–300. Available: http://www.nature.com/doifinder/10.1038/485298a. Accessed 16 May 2012.

33. Makel MC, Plucker JA, Hegarty B (2012) Replications in Psychology Research: How Often Do They Really Occur? Perspectives on Psychological Science 7: 537–542. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691612460688. Accessed 8 November 2012.

34. Editorial (2012) The Well-Behaved Scientist. Science 335: 285–285. Available: http://www.sciencemag.org/cgi/doi/10.1126/science.335.6066.285-b. Accessed 8 March 2012.

35. Goldacre B (2011) I foresee that nobody will do anything about this problem – Bad Science. Bad Science. Available: http://www.badscience.net/2011/04/i-foresee-that-nobody-will-do-anything-about-this-problem/. Accessed 8 March 2012.

36. Sutton J (2011) psi study highlights replication problems. The Psychologist News. Available: http://www.thepsychologist.org.uk/blog/blogpost.cfm?threadid=1984&catid=48. Accessed 8 March 2012.

37. Curry S (2009) Eye-opening Access. Occam's Typwriter: Reciprocal Space. Available: http://occamstypewriter.org/scurry/2009/03/27/eye_opening_access/.

38. Baker M (2012) Independent labs to verify high-profile papers. Nature. Available: http://www.nature.com/doifinder/10.1038/nature.2012.11176. Accessed 8 January 2013.

39. Collaboration OS (2012) An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. Perspectives on Psychological Science 7: 657–660. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691612462588. Accessed 8 November 2012.

40. Gordon MD (1982) Citation ranking versus subjective evaluation in the determination of journal hierachies in the social sciences. Journal of the American Society for Information Science 33: 55–57. doi:10.1002/asi.4630330109.

41. Saha S, Saint S, Christakis DA (2003) Impact factor: a valid measure of journal quality? Journal of the Medical Library Association : JMLA 91: 42–46. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=141186&tool=pmcentrez&rendertype=abstract. Accessed 21 March 2012.

42. Yue W, Wilson CS, Boller F (2007) Peer assessment of journal quality in clinical neurology. Journal of the Medical Library Association : JMLA 95: 70–76. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1773051&tool=pmcentrez&rendertype=abstract. Accessed 17 December 2012.

43. Sønderstrup-Andersen EM, Sønderstrup-Andersen HHK (2008) An investigation into diabetes researcher's perceptions of the Journal Impact Factor — reconsidering evaluating research. Scientometrics 76: 391–406. Available: http://www.springerlink.com/index/10.1007/s11192-007-1924-4. Accessed 15 November 2012.

44. Steen RG (2011) Retractions in the scientific literature: do authors deliberately commit research fraud? Journal of medical ethics 37: 113–117. doi:10.1136/jme.2010.038125.

45. Munafò MR, Freimer NB, Ng W, Ophoff R, Veijola J, et al. (2009) 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data.

American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 150B: 271–281. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2819421&tool=pmcentrez&rendertype=abstract. Accessed 14 March 2012.

46. Siontis KCM, Evangelou E, Ioannidis JPA (2011) Magnitude of effects in clinical trials published in high-impact general medical journals. International journal of epidemiology 40: 1280–1291. Available: http://www.ncbi.nlm.nih.gov/pubmed/22039194. Accessed 8 March 2012.

47. Ioannidis JPA (2005) Contradicted and initially stronger effects in highly cited clinical research. JAMA : the journal of the American Medical Association 294: 218–228. Available: http://www.ncbi.nlm.nih.gov/pubmed/16014596. Accessed 4 March 2012.

48. Ioannidis JPA, Panagiotou OA (2011) Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA : the journal of the American Medical Association 305: 2200–2210. Available: http://www.ncbi.nlm.nih.gov/pubmed/21632484. Accessed 10 March 2012.

49. Fang FC, Casadevall A (2011) Retracted science and the retraction index. Infection and immunity 79: 3855–3859. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3187237&tool=pmcentrez&rendertype=abstract. Accessed 7 March 2012.

50. Liu S (2006) Top Journals' Top Retraction Rates. Scientific Ethics 1: 92–93.

51. Anderson MS, Martinson BC, De Vries R (2007) Normative dissonance in science: results from a national survey of u.s. Scientists. Journal of empirical research on human research ethics : JERHRE 2: 3–14. Available: http://www.ncbi.nlm.nih.gov/pubmed/19385804. Accessed 21 March 2012.

52. Shapin S (2008) The scientific life : a moral history of a late modern vocation. Chicago: University of Chicago Press.

53. Fanelli D (2011) Negative results are disappearing from most disciplines and countries. Scientometrics 90: 891–904. Available: http://www.springerlink.com/content/bt125xw73v72282w/. Accessed 8 March 2012.

54. Giles J (2007) Breeding cheats. Nature 445: 242–243. Available: http://www.ncbi.nlm.nih.gov/pubmed/17230160. Accessed 21 March 2012.

55. Hegarty P, Walton Z (2012) The Consequences of Predicting Scientific Impact in Psychology Using Journal Impact Factors. Perspectives on Psychological Science 7: 72–78. Available: http://pps.sagepub.com/lookup/doi/10.1177/1745691611429356. Accessed 8 March 2012.

56. Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. BMJ 314. Available: http://www.bmj.com/cgi/content/abstract/314/7079/497.

57. Kravitz DJ, Baker CI (2011) Toward a new model of scientific publishing: discussion and a proposal. Frontiers in computational neuroscience 5: 55. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3230039&tool=pmcentrez&rendertype=abstract. Accessed 2 March 2012.

58. Chow CW, Haddad K, Singh G, Wu A (2007) On Using Journal Rank to Proxy for an Article ' s Contribution or Value. Issues in Accounting Education 22: 411–427.

59. Seglen PO (1994) Causal relationship between article citedness and journal impact. Journal of the American Society for Information Science 45: 1–11. Available: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y/abstract. Accessed 21 March 2012.

60. Lozano GA, Lariviere V, Gingras Y (2012) The weakening relationship between the Impact Factor and papers' citations in the digital age. ArXiv: 14. Available: http://arxiv.org/abs/1205.4328. Accessed 29 May 2012.

61. Singh G, Haddad KM, Chow CW (2007) Are Articles in "Top" Management Journals Necessarily of Higher Quality? Journal of Management Inquiry 16: 319–331. Available: http://jmi.sagepub.com/cgi/doi/10.1177/1056492607305894. Accessed 17 December 2012.

62. Starbuck WH (2005) How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication. Organization Science 16: 180–200. Available: http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1040.0107. Accessed 17 December 2012.

63. Weale AR, Bailey M, Lear PA (2004) The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. BMC medical research methodology 4: 14. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=434502&tool=pmcentrez&rendertype=abstract. Accessed 2 April 2012.

64. Google (n.d.) Google Scholar Citations. Available: http://scholar.google.com/citations?view_op=top_venues&hl=en&vq=en. Accessed 2 April 2012.

65. Allen L, Jones C, Dolby K, Lynn D, Walport M (2009) Looking for Landmarks: The Role of Expert Review and Bibliometric Analysis in Evaluating Scientific Publication Outputs. PLoS ONE 4: 8. Available: http://dx.plos.org/10.1371/journal.pone.0005910.

66. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P, Bhandari M (2005) Level of evidence in orthopaedic journals. The Journal of bone and joint surgery American volume 87: 2632–2638. Available: http://www.ncbi.nlm.nih.gov/pubmed/16322612. Accessed 17 December 2012.

67. Lau SL, Samman N (2007) Levels of evidence and journal impact factor in oral and maxillofacial surgery. International journal of oral and maxillofacial surgery 36: 1–5. Available: http://www.ncbi.nlm.nih.gov/pubmed/17129707. Accessed 17 December 2012.

68. Bain CR, Myles PS (2005) Relationship between journal impact factor and levels of evidence in anaesthesia. Anaesthesia and intensive care 33: 567–570. Available: http://www.ncbi.nlm.nih.gov/pubmed/16235472.

69. Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, et al. (2008) Design, power, and interpretation of studies in the standard murine model of ALS. Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases 9: 4–15. Available: http://www.ncbi.nlm.nih.gov/pubmed/18273714. Accessed 23 March 2012.

70. Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. Nature 483: 531–533. Available: http://www.nature.com/doifinder/10.1038/483531a. Accessed 29 March 2012.

71. Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nature reviews Drug discovery 10: 712. Available: http://www.ncbi.nlm.nih.gov/pubmed/21892149. Accessed 10 March 2012.

72. Statzner B, Resh VH (2010) Negative changes in the scientific publication process in ecology: potential causes and consequences. Freshwater Biology 55: 2639–2653. Available: http://doi.wiley.com/10.1111/j.1365-2427.2010.02484.x. Accessed 11 March 2012.

73. Nosek BA, Bar-Anan Y (2012) Scientific Utopia: I. Opening Scientific Communication. Psychological Inquiry 23: 217–243. Available: http://www.tandfonline.com/doi/abs/10.1080/1047840X.2012.692215. Accessed 8 November 2012.

74. Calcagno V, Demoinet E, Gollner K, Guidi L, Ruths D, et al. (2012) Flows of Research Manuscripts Among Scientific Journals Reveal Hidden Submission Patterns. Science (New York, NY) 338: 1065–1069. Available: http://www.sciencemag.org/content/338/6110/1065.abstract. Accessed 26 October 2012.

75. Lu H, Zou Q, Gu H, Raichle ME, Stein EA, et al. (2012) Rat brains also have a default mode network. Proceedings of the National Academy of Sciences of the United States of America 109: 3979–3984. Available: http://www.ncbi.nlm.nih.gov/pubmed/22355129. Accessed 9 March 2012.

76. Upadhyay J, Baker SJ, Chandran P, Miller L, Lee Y, et al. (2011) Default-mode-like network activation in awake rodents. PloS one 6: e27839. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3220684&tool=pmcentrez&rendertype=abstract. Accessed 8 March 2012.

77. Welberg L (2012) Neuroimaging: Rats join the “default mode” club. Nature reviews Neuroscience. Available: http://www.ncbi.nlm.nih.gov/pubmed/22414942. Accessed 18 March 2012.

78. Kyrillidou M, Morris S, Roebuck G (2012) ARL statistics. American Research Libraries Digital Publications. Available: http://publications.arl.org/ARL_Statistics. Accessed 18 March 2012.

79. House of Commons (2004) House of Commons - Science and Technology - Written Evidence. Parliamentary Publications and Records. Available: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399we163.htm. Accessed 17 December 2012.

80. Editorial (2006) The impact factor game. It is time to find a better way to assess the scientific literature. PLoS medicine 3: e291. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1475651&tool=pmcentrez&rendertype=abstract. Accessed 7 March 2012.

81. Editorial (2005) Not-so-deep impact. Nature 435: 1003–1004. Available: http://www.ncbi.nlm.nih.gov/pubmed/15973363. Accessed 14 March 2012.

82. Garfield E (1999) Journal impact factor: a brief review. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 161: 979–980. Available: http://www.cmaj.ca/content/161/8/979.full. Accessed 14 March 2012.

83. Adam D (2002) The counting house. Nature 415: 726–729. Available: http://www.ncbi.nlm.nih.gov/pubmed/11845174. Accessed 21 March 2012.

84. Moed HF, Van Leeuwen TN (1995) Improving the accuracy of institute for scientific information's journal impact factors. Journal of the American Society for Information Science 46: 461–467. Available: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199507)46:6<461::AID-ASI5>3.0.CO;2-G/abstract. Accessed 21 March 2012.

85. Moed HF, Van Leeuwen TN (1996) Impact factors can mislead. Nature 381: 186. Available: http://dx.doi.org/10.1038/381186a0. Accessed 16 April 2012.

86. Rossner M, Van Epps H, Hill E (2007) Show me the data. The Journal of Cell Biology 179: 1091–1092. Available: http://jcb.rupress.org/content/179/6/1091.abstract.

87. Reedijk J (1998) Sense and nonsense of science citation analyses: comments on the monopoly position of ISI and citation inaccuracies. Risks of possible misuse and biased citation and impact data. New Journal of Chemistry 22: 767–770. doi:10.1039/A802808G.

88. Moed HF, Van Leeuwen TN, Reedijk J (1996) A critical analysis of the journal impact factors ofAngewandte Chemie and the journal of the American Chemical Society inaccuracies in published impact factors based on overall citations only. Scientometrics 37: 105–116. doi:10.1007/BF02093487.

89. Seglen PO (1992) The skewness of science. Journal of the American Society for Information Science 43: 628–638. Available: http://doi.wiley.com/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0. Accessed 14 March 2012.

90. Taylor M, Perakakis P, Trachana V (2008) The siege of science. Ethics in Science and Environmental Politics 8: 17–40. Available: http://www.int-res.com/abstracts/esep/v8/n1/p17-40/. Accessed 5 December 2012.

91. Adler R, Ewing J, Taylor P (2008) Joint Committee on Quantitative Assessment of Research: Citation Statistics (A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathemat. Available: http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf.

92. Vanclay JK (2011) Impact factor: outdated artefact or stepping-stone to journal certification? Scientometrics 92: 211–238. Available: http://www.springerlink.com/index/10.1007/s11192-011-0561-0. Accessed 9 November 2012.

93. Wagner PD (2011) What's in a number? Journal of applied physiology (Bethesda, Md : 1985). Available: http://www.ncbi.nlm.nih.gov/pubmed/21799127. Accessed 24 March 2012.

94. Hauser JR, Katz GM (1998) Metrics: you are what you measure! European Management Journal 16: 517–528. Available: http://linkinghub.elsevier.com/retrieve/pii/S0263237398000292.

95. Frank M (2003) Impact factors: arbiter of excellence? Journal of the Medical Library Association : JMLA 91: 4–6. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=141180&tool=pmcentrez&rendertype=abstract. Accessed 16 April 2012.

96. Adler NJ, Harzing a.-W (2009) When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings. Academy of Management Learning & Education 8: 72–95. Available: http://amle.aom.org/cgi/doi/10.5465/AMLE.2009.37012181.

97. Bauer HH (2004) Science in the 21st Century : Knowledge Monopolies and Research Cartels. Jour Scient Explor 18: 643–660.

98. Lawrence PA (2002) Rank injustice. Nature 415: 835–836. Available: http://www.ncbi.nlm.nih.gov/pubmed/11859339. Accessed 17 December 2012.

99. Brumback RA (2012) "3 . . 2 . . 1 . . Impact [factor]: target [academic career] destroyed!": just another statistical casualty. Journal of child neurology 27: 1565–1576. Available: http://www.ncbi.nlm.nih.gov/pubmed/23183597. Accessed 17 December 2012.

100. Lawrence PA (2007) The mismeasurement of science. Current biology : CB 17: R583–5. Available: http://www.ncbi.nlm.nih.gov/pubmed/17686424. Accessed 5 November 2012.

101. Lawrence P (2008) Lost in publication: how measurement harms science. Ethics in Science and Environmental Politics 8: 9–11. Available: http://www.int-res.com/abstracts/esep/v8/n1/p9-11/. Accessed 28 October 2012.

102. Garwood J (2011) A conversation with Peter Lawrence, Cambridge. “The Heart of Research is Sick”. LabTimes 2-2011: 24–31. Available: http://www.labtimes.org/labtimes/issues/lt2011/lt02/lt_2011_02_24_31.pdf. Accessed 17 December 2012.

103. Tsikliras A (2008) Chasing after the high impact. Ethics in Science and Environmental Politics 8: 45–47. Available: http://www.int-res.com/abstracts/esep/v8/n1/p45-47/. Accessed 5 December 2012.

104. Todd P, Ladle R (2008) Hidden dangers of a □citation cultu Environmental Politics 8: 13–16. Available: http://www.int-res.com/abstracts/esep/v8/n1/p13-16/. Accessed 5 November 2012.

105. Sarewitz D (2012) Beware the creeping cracks of bias. Nature 485: 149–149. Available: http://www.nature.com/doifinder/10.1038/485149a. Accessed 10 May 2012.

106. Gauchat G (2010) The cultural authority of science: Public trust and acceptance of organized science. Public Understanding of Science 20: 751–770. Available: http://pus.sagepub.com/cgi/doi/10.1177/0963662510365246. Accessed 6 March 2012.

107. EuropeanCommission (2010) Science and Technology Report.

108. Nowotny H (2005) Science and society. High- and low-cost realities for science and society. Science (New York, NY) 308: 1117–1118. Available: http://www.ncbi.nlm.nih.gov/pubmed/15905384. Accessed 8 March 2012.

109. Gauchat G (2012) Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010. American Sociological Review 77: 167–187. Available: http://asr.sagepub.com/cgi/doi/10.1177/0003122412438225. Accessed 29 March 2012.

110. Miller KR (2012) America’s Darwin Problem. Huffington Post. Available: http://www.huffingtonpost.com/kenneth-r-miller/darwin-day-evolution_b_1269191.html. Accessed 14 March 2012.

111. Sarewitz D (2013) Science must be seen to bridge the political divide. Nature 493: 7. Available: http://www.ncbi.nlm.nih.gov/pubmed/23282339. Accessed 8 January 2013.

112. Lin J (2012) Cracking Open the Scientific Process: “Open Science” Challenges Journal Tradition With Web Collaboration. New York Times. Available: http://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html?pagewanted=all. Accessed 8 March 2012.

113. Priem J, Piwowar HA, Hemminger BM (2012) Altmetrics in the wild: Using social media to explore scholarly impact. ArXiv 1203.4745. Available: http://arxiv.org/abs/1203.4745. Accessed 13 April 2012.

114. Hönekopp J, Khan J (2011) Future publication success in science is better predicted by traditional measures than by the h index. Scientometrics 90: 843–853. Available:

http://www.springerlink.com/index/10.1007/s11192-011-0551-2. Accessed 7 November 2012.

115. Roemer RC, Borchardt R (2012) From bibliometrics to altmetrics: A changing scholarly landscape. College & Research Libraries News 73: 596–600. Available: http://crln.acrl.org/cgi/content/abstract/73/10/596.

116. Eve MP (2012) Tear it down, build it up: the Research Output Team, or the library-as-publisher. Insights: the UKSG journal 25: 158–162. Available: http://uksg.metapress.com/openurl.asp?genre=article&id=doi:10.1629/2048-7754.25.2.158. Accessed 10 July 2012.

117. Beverungen A, Bohm S, Land C (2012) The poverty of journal publishing. Organization 19: 929–938. Available: http://org.sagepub.com/cgi/doi/10.1177/1350508412448858. Accessed 3 November 2012.

118. Casadevall A, Fang FC (2012) Reforming science: methodological and cultural reforms. Infection and immunity 80: 891–896. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3294645&tool=pmcentrez&rendertype=abstract. Accessed 27 March 2012.

119. Andrew T (2012) Gold Open Access: Counting the Costs | Ariadne. Ariadne (70) 70. Available: http://www.ariadne.ac.uk/issue70/andrew. Accessed 23 December 2012.

120. Houghton J, Swan A (2013) Planting the Green Seeds for a Golden Harvest: Comments and Clarifications on “Going for Gold”. D-Lib Magazine 19. Available: http://www.dlib.org/dlib/january13/houghton/01houghton.html. Accessed 16 January 2013.

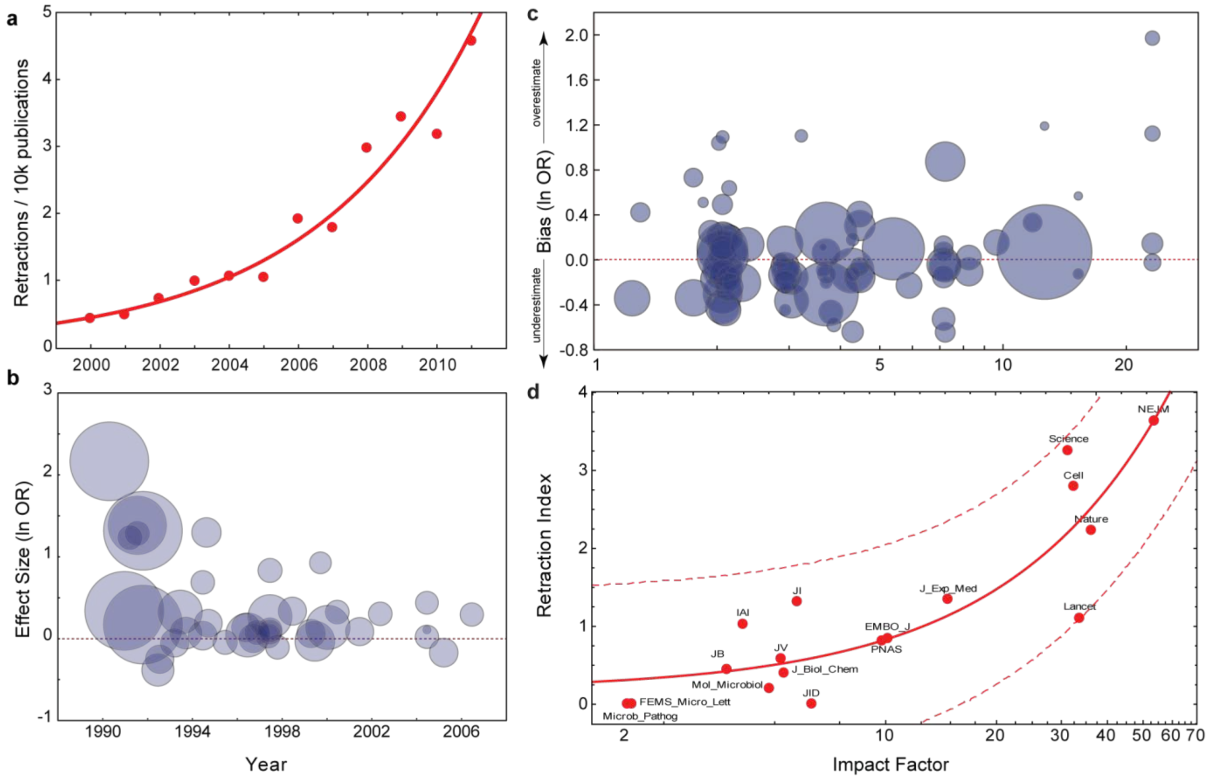
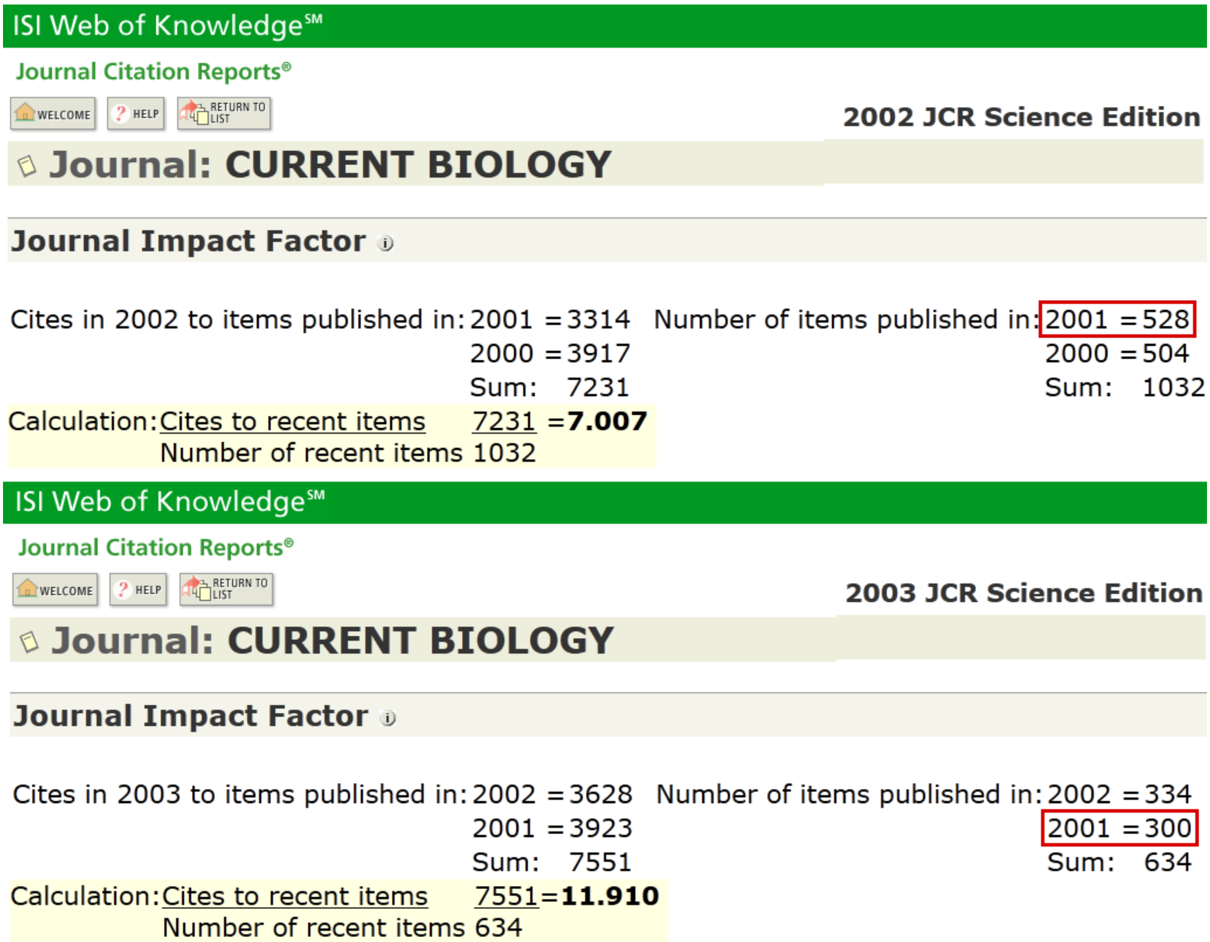



**Fig. 1:** *Current trends in the reliability of science.*

**a** – Exponential fit for PubMed retraction notices (data from pmretract.heroku.com). **b** – Relationship between year of publication and individual study effect size (circle area is proportional to IF; data taken from Munafò et al., 2007). **c** – Relationship between IF and extent to which an individual study overestimates the likely true effect. Both effect size overestimation and reduced sample size are significantly correlated with IF (circle area is proportional to sample size; data from Munafò et al., 2009). **d** – Linear regression with confidence intervals between IF and Fang and Casadevall's Retraction Index (data provided by Fang and Casadevall, 2011).

| Journal: Current Biology | Published items 2000 | Published items 2001 | Published items 2002 | Sum published items | Citations in preceding two years | IF |
|---|---|---|---|---|---|---|
| JCR Science Edition 2002 | 504 | 528 | n.c. | 1032 | 7231 | 7.007 |
| JCR Science Edition 2003 | n.c. | 300 | 334 | 634 | 7551 | 11.910 |

**Table 1:** Thomson Reuters' IF calculations for the journal 'Current Biology' in the years 2002/2003. Most of the rise in IF is due to the reduction in published items. Note the discrepancy between the number of items published in 2001 between the two consecutive JCR Science Editions. – n.c.: year not covered by this edition. Raw data see Suppl. Fig. S1.



ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME | HELP | RETURN TO LIST

2002 JCR Science Edition

Journal: CURRENT BIOLOGY

Journal Impact Factor

| | | | |
|---|---|---|---|
| Cites in 2002 to items published in: | 2001 = 3314 | Number of items published in: | 2001 = 528 |
| | 2000 = 3917 | | 2000 = 504 |
| | Sum: 7231 | | Sum: 1032 |

Calculation: Cites to recent items / Number of recent items = 7231 / 1032 = **7.007**

ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME | HELP | RETURN TO LIST

2003 JCR Science Edition

Journal: CURRENT BIOLOGY

Journal Impact Factor

| | | | |
|---|---|---|---|
| Cites in 2003 to items published in: | 2002 = 3628 | Number of items published in: | 2002 = 334 |
| | 2001 = 3923 | | 2001 = 300 |
| | Sum: 7551 | | Sum: 634 |

Calculation: Cites to recent items / Number of recent items = 7551 / 634 = **11.910**

**Suppl. Fig. S1:** *Impact Factor of the journal „Current Biology" in the years 2002 (above) and 2003 (below) showing a 40% increase in impact.*

The increase in the IF of the journal "Current Biology" from approx. 7 to almost 12 from one edition of Thomson Reuters' "Journal Citation Reports" to the next is due to a retrospective adjustment of the number of items published (marked), while the actual citations remained relatively constant.