



# Prestigious Science Journals Struggle to Reach Even Average Reliability

Björn Brembs\*

*Institute of Zoology—Neurogenetics, Universität Regensburg, Regensburg, Germany*

In which journal a scientist publishes is considered one of the most crucial factors determining their career. The underlying common assumption is that only the best scientists manage to publish in a highly selective tier of the most prestigious journals. However, data from several lines of evidence suggest that the methodological quality of scientific experiments does not increase with increasing rank of the journal. On the contrary, an accumulating body of evidence suggests the inverse: methodological quality and, consequently, reliability of published research works in several fields may be *decreasing* with increasing journal rank. The data supporting these conclusions circumvent confounding factors such as increased readership and scrutiny for these journals, focusing instead on quantifiable indicators of methodological soundness in the published literature, relying on, in part, semi-automated data extraction from often thousands of publications at a time. With the accumulating evidence over the last decade grew the realization that the very existence of scholarly journals, due to their inherent hierarchy, constitutes one of the major threats to publicly funded science: hiring, promoting and funding scientists who publish unreliable science eventually erodes public trust in science.

## OPEN ACCESS

### Edited by:

Hauke R. Heekeren,  
Freie Universität Berlin, Germany

### Reviewed by:

Richard C. Gerkin,  
Arizona State University,  
United States  
Andreas Prokop,  
University of Manchester,  
United Kingdom

### \*Correspondence:

Björn Brembs  
bjoem@brembs.net

**Received:** 09 November 2017

**Accepted:** 23 January 2018

**Published:** 20 February 2018

### Citation:

Brembs B (2018) Prestigious Science Journals Struggle to Reach Even Average Reliability. *Front. Hum. Neurosci.* 12:37. doi: 10.3389/fnhum.2018.00037

**Keywords:** journals, journal ranking, reliability, reproducibility of results, science policy

## INTRODUCTION

The most groundbreaking, transformative research results deserve a broad readership and a large audience. Therefore, scientists submit their best work to the journals with the largest audience. While the number of scientists has been growing exponentially over the last decades, the number of journals with a large audience has not kept up, neither has the number of articles published per journal. Consequently, rejection rates at the most prestigious journals has fallen below 10% and the labor of rejecting submissions has become these journals' largest cost item. Assuming that this exclusivity allows the journals to separate the wheat from the chaff, successful publication in these journals is treated as a quality signal in hiring, promotion and funding decisions. If anything, these developments have fueled the circularity of this relationship: today, publishing ground-breaking science in a high ranking journals is not only important for science to advance but also for an author's career to advance. Even before science became hypercompetitive at every level, now and again results published in prestigious journals were later found to be false. This is the nature of science. Science is difficult, complicated and perpetually preliminary. Science is self-correcting and better experimentation will continue to advance science to the detriment of previous experiments. Today, however, fierce competition exacerbates this trait and renders it a massive problem for scholarly journals.

Now it has become their task to find the ground-breaking among the too-good-to-be-true data, submitted by desperate scientists, who face unemployment and/or laboratory closure without the next high-profile publication. This is a monumental task, given that sometimes it takes decades to find that one or the other result rests on flimsy grounds. How is our hierarchy of more than 30,000 journals holding up?

At first glance, it appears as if our journals fail miserably. Evaluating retractions, the capital punishment for articles found to be irreproducible, it was found that the most prestigious journals boast the largest number (Fang and Casadevall, 2011) and that most of these retractions are due to fraud (Fang et al., 2012). However, data on retractions suffer from two major flaws which make them rather useless for answering questions about the contribution of journals to the reliability (or lack thereof) of our scholarly literature: (1) retractions cover only about 0.05% of the literature; and (2) they are confounded by error-detection variables that are hard to trace. So may be our journals are not doing so horribly after all?

## JOURNAL RANKING

The most widely used metric to rank journals is Clarivate Analytics' "Impact Factor" (IF), a measure based loosely on citations. Despite the numerous flaws described (e.g., Moed and van Leeuwen, 1996; Seglen, 1997; Saha et al., 2003; Rossner et al., 2007; Adler et al., 2009; Hernán, 2009; Vanclay, 2011; Brembs et al., 2013), the IF is an excellent and consistent descriptor of subjective journal hierarchy, i.e., the level of prestige scientists ascribe to the journals in their respective fields (Gordon, 1982; Saha et al., 2003; Yue et al., 2007; Sønderstrup-Andersen and Sønderstrup-Andersen, 2008). That a measure so flawed still conforms to the expectations of the customers expected to pay for it, is remarkable in its own right. Due to this consistency, the IF is used here as a measure for the subjective ranking of journals by the scientists using these journals: to what extent is this subjective notion of prestige warranted, based on the available evidence? Are prestigious journals really better at detecting the real breakthrough science in the sea of seemingly breakthrough science than average journals?

## RETRACEMENTS AND ERROR DETECTION

If anything, one could tentatively interpret what scant data there are on retractions, as suggestive that increased scrutiny may only play a minor role in a combination of several factors leading to more retractions in higher ranking journals.

For instance, there are low ranking journals with high retraction rates (Fang et al., 2012), showing that the involved parties are motivated to retract articles even in low ranking journals. In fact, in absolute terms, most retracted articles come from low-ranking journals. This would be difficult to explain if low ranking journals were less willing to retract and/or scholars less motivated to pursue retractions from these journals. On the other hand, one can make the claim that the numbers

would show even more retractions in low ranking journals, if the motivation and willingness to retract were equal. As neither willingness of journals to retract nor motivation in individuals to force a retraction can be quantified, all that the data can show is that it is not an all-or-nothing effect: there is both willingness and motivation to retract also for articles in lower ranking journals.

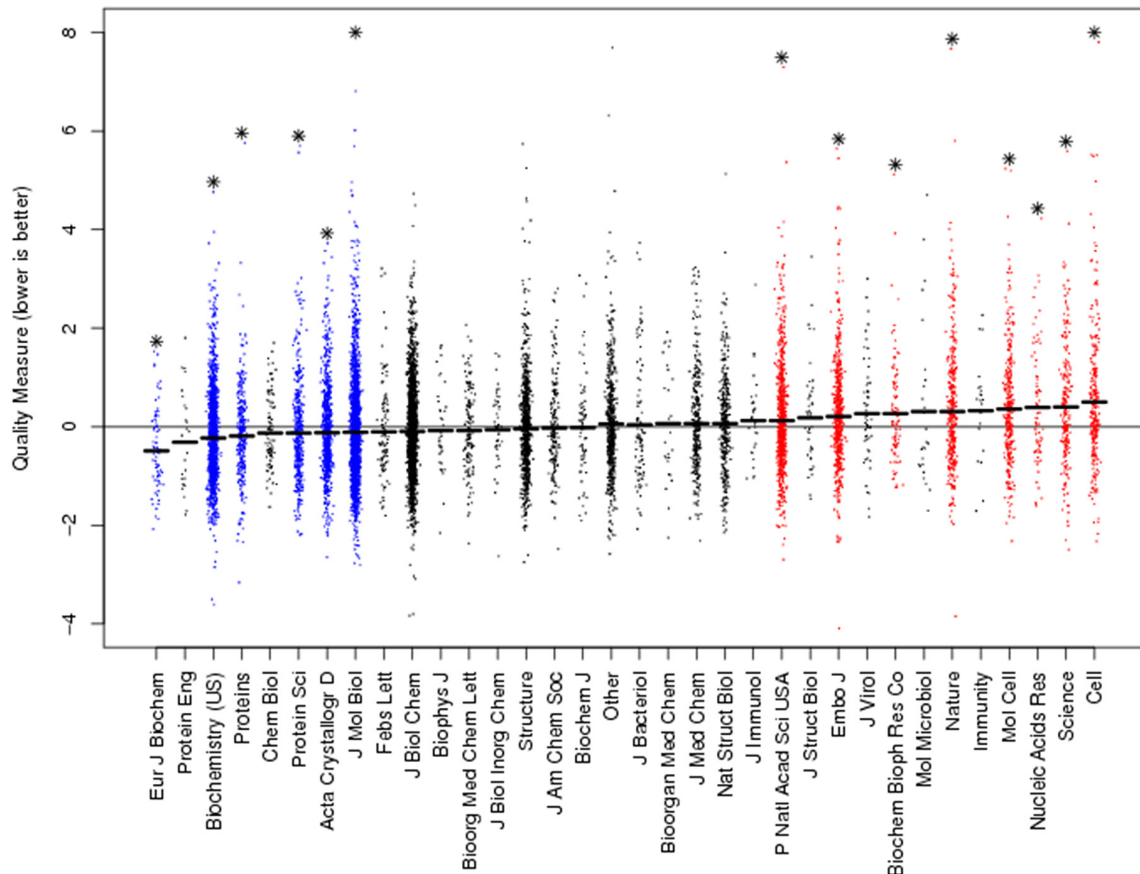
Another reason why scrutiny might be assumed to be higher in more prestigious journals is that readership is higher, leading to more potential for error detection. More eyes are more likely to detect potential errors. The consequence of this reasonable and plausible factor is difficult to test empirically. However, one could make a more easily testable, analogous claim, such as that one would also expect increased readership to lead to a higher potential not only for retractions but also for citations. More eyes are more likely to detect a finding worth citing. In fact, if anything, citations ought to correlate better with journal rank than retractions because citing an article in a leading journal is not only technically easier than forcing a retraction, it also benefits one's own research by elevating the perceived importance of one's own field. However, the opposite is the case: The coefficient of determination for citations with journal rank currently lies around 0.2, while that coefficient comes to lie at just under 0.8 for retractions and journal rank (Brembs et al., 2013). So while there may be a small effect of scrutiny/motivation, the evidence seems to suggest that it is a relatively minor effect, if there is one at all.

Taken together, there is currently no strong case to be made as to whether the likely increased scrutiny and readership of highly-ranked journals is a major factor driving retractions or not. If that were the case, it would indicate that the apparent increased unreliability in high-ranking journals is merely an artifact of the increased scrutiny to retract, combined with an increased willingness of these journals to correct the scientific record. At least two lines of inquiry did not turn up any conclusive evidence for such an argument. With such unclarified confounds in such a tiny section of the literature, it is straightforward to disregard retractions as extreme outliers and focus instead of the 99.95% of unretracted articles in order to estimate the reliability of highly ranked journals.

## THE OTHER 99.95% OF THE LITERATURE

In the literature covering unretracted, peer-reviewed articles, one can identify at least eight lines of evidence suggesting that articles published in higher ranking journals are methodologically either not stronger or, indeed, weaker than those in lower ranking journals. In contrast, there is no evidence that articles published in higher ranking journals are methodologically stronger. Methodology here refers to several measures of experimental and statistical rigor with a potential bearing on subsequent replication or re-use. There is currently one article with evidence that higher ranking journals are better at detecting duplicated images (Bik et al., 2016).

In the following, I will quickly review the lines of evidence in the order of decreasing evidential strength.



**FIGURE 1** | Ranking journals according to crystallographic quality reveals high-ranking journals with the lowest quality work. The quality metric (y-axis) is computed as a deviation from perfect. Hence, lower values denote higher quality work. Each dot denotes a single structure. The quality metric was normalized to the sample average and journals ranked according to their mean quality. Asterisks denote significant difference from sample average. Figure courtesy of Dr. Ramaswamy, methods in Brown and Ramaswamy (2007).

## Crystallographic Quality

The quality of computer models of molecular structures, derived from crystallographic work, can be quantified by a method which includes the deviations from known atomic distances and other factors (Brown and Ramaswamy, 2007). Averaging the quality metric for each journal, high-ranking journals such as *Cell*, *Molecular Cell*, *Nature*, *EMBO Journal* and *Science* publish significantly substandard structures (Figure 1, courtesy of Dr. Ramaswamy, methods in Brown and Ramaswamy, 2007). The molecular complexity or the difficulty of the crystallographic work cannot explain this finding, as these factors are incorporated in the computation of the quality metric.

## Effect Sizes in Gene-Association Studies

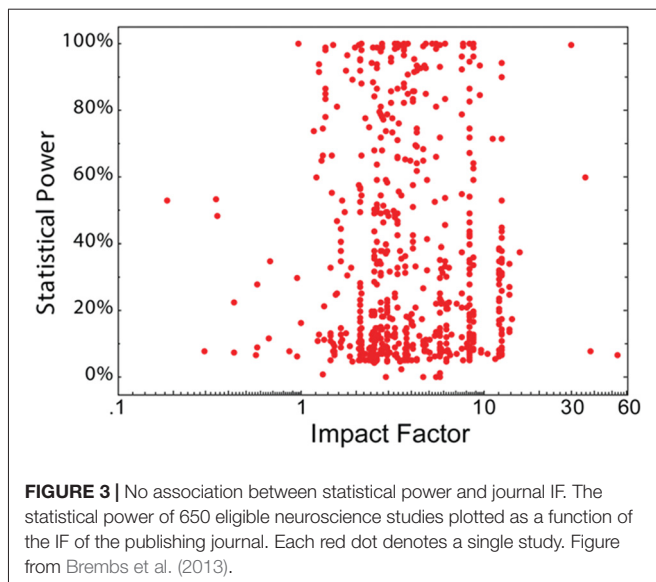
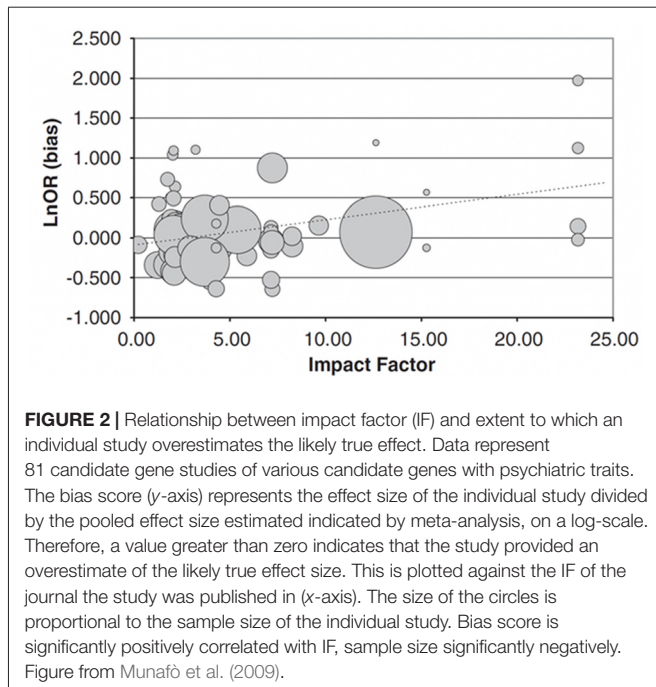
Analyzing effect sizes in gene-association studies, Munafò et al. (2009) found that for 81 different studies on psychiatric traits, higher ranking journals overestimated the size of the gene-trait association, while the sample size decreased with increasing ranking of the journal (Figure 2). Phrased differently and more generally, inflated effect sizes are disproportionately often found

in journals which rank more highly and publish studies with lower sample sizes.

## Statistical Power in Neuroscience/Psychology

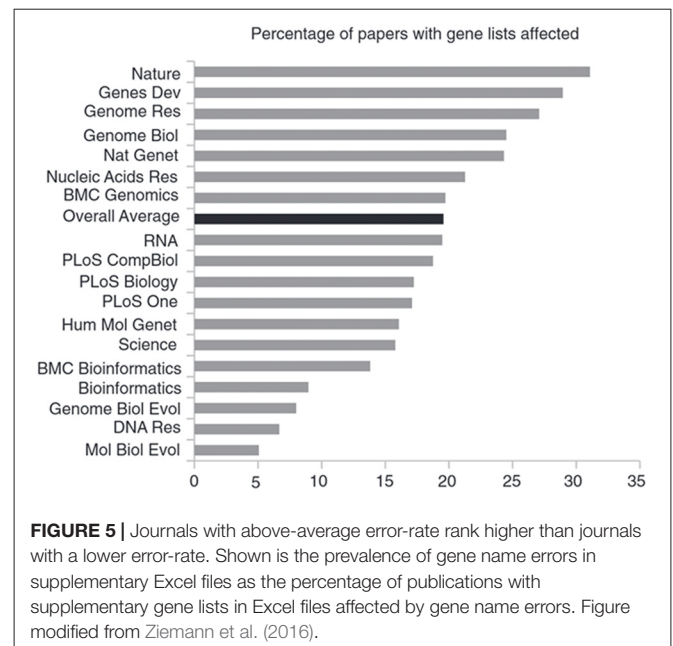
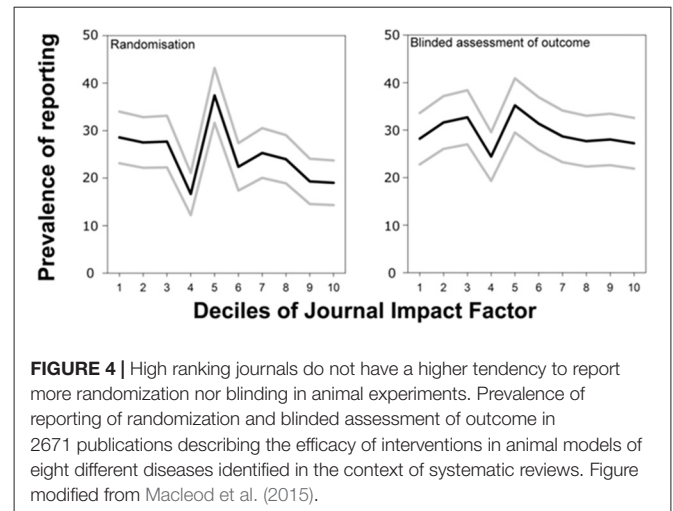
Statistical power (defined as  $1 - \text{type II error rate}$ ; a measure computed from sample size and effect size) allows inference as to the likelihood that a nominally statistically significant finding actually reflects a true effect. As such, statistical power is directly related to the reliability of the experiments conducted. Button et al. (2013) analyzed the statistical power of 730 individual primary neuroscience studies. These data do not show any correlation with journal rank (Brembs et al., 2013; Figure 3).

Cognitive neuroscience and psychology also seem to suffer from insufficient statistical power. In their case, unlike in the neuroscience case, statistical power is even *negatively* correlated with journal rank (Szucs and Ioannidis, 2017). Taken together, these results suggest that in the covered fields at least, results from higher ranking journals tend to be less reliable than those from lower ranking journals, with a low overall reliability, as expressed in statistical power.



## Experimental Design in *in Vivo* Animal Experimentation

In preclinical research studying animal models of disease, a widely used standard experimental design requires randomized assortment of animals into the treatment and control group, respectively, as well as an outcome assessment where the assessor scoring the outcome is blind as to the treatment group of the animal. Analyzing the methods sections of publications reporting *in vivo* experiments of animal disease models where this design should have been applied, Macleod et al. (2015) found that the prevalence of reporting randomization before the treatment correlated negatively with journal rank, while



reporting of blind outcome assessment was not correlated (Figure 4, modified from Macleod et al., 2015). Inasmuch as this reporting correlates with actual experimentation, such publications in higher ranking journals would hence be less reliable than those in other journals: not reporting bias precludes replication and a reported bias may still entail that the bias created the observed effect. If randomization were not rarer (or even more frequent) in high vs. low-ranking journals, authors of articles in high-ranking journals are at least failing to report this randomization more often than authors in lower ranking journals.

## Errors in Genomics, Cognitive Neuroscience and Psychology

With the advent of data deposition mandates by some journals, it has become a regular practice to supply the required data as supplemental files, often in spreadsheet form,

compatible with the Excel software (Microsoft Corp., Redmond, WA, USA). However, when using Excel's default settings, gene symbols and accession numbers may inadvertently be converted into dates or floating point numbers (Ziemann et al., 2016). These errors are widespread in the literature and the journals with above average error-rate are more highly ranked than journals with below-average error rate (Figure 5, modified from Ziemann et al., 2016). The authors speculate that the correlation they found is due to higher ranking journals publishing larger gene collections. This explanation, if correct, would suggest that, on average, error detection in such journals is at least not superior to that in other journals.

Another source of error in the literature are *p*-value reporting errors, i.e., the *p*-values reported in a publication deviate from the *p*-value calculated from the data. Comparing only those *p*-value reporting errors that changed the significance of the outcome across their sample of 18 journals in cognitive neuroscience and psychology, Szucs and Ioannidis (2016) found a significant correlation between the rate of such erroneous records and the IF of the journal (Figure 6, redrawn from Szucs and Ioannidis, 2016). Interestingly, the errors were highly skewed in the direction of reporting a non-significant computed *p*-value as significant (Szucs and Ioannidis, 2016).

## Criteria for Evidence-Based Medicine

One way of estimating the reliability of scientific publications is to count how many criteria for evidence-based medicine (e.g., according to JBJS-A, Australian National Health and Medical Research Council guidelines or GRADE) have been fulfilled and correlate these data with journal rank. There currently are four such studies, where two found such a correlation, indicating higher-ranked journals publish work where more criteria are fulfilled (Obremskey et al., 2005; Lau and Samman, 2007) and two did not (Bain and Myles, 2005; Tressoldi et al., 2013). In the case of these studies, if there is any actual correlation between journal rank and levels of evidence in medicine, it appears to be

too weak to be detected consistently. For further discussion of these studies see Brembs et al. (2013).

## Reliability Metrics in Psychology

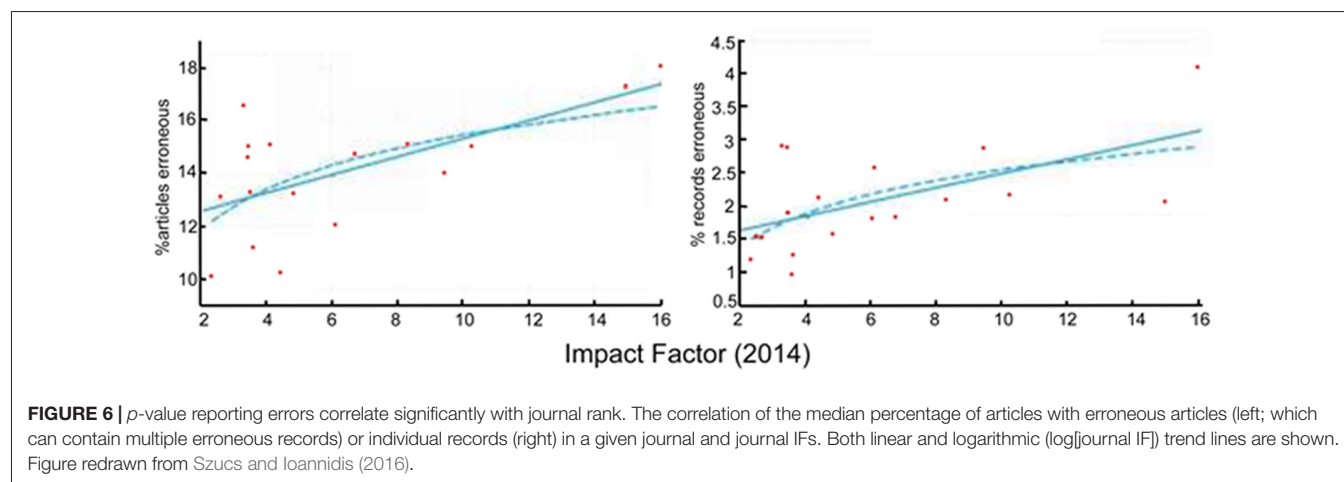
In the wake of recent replication efforts on psychology (Open Science Collaboration, 2015), metrics have been developed to detect tell-tale signs of questionable research practices or publication bias in a body of published work. These metrics commonly use *p*-value distributions or statistical power and other values which can be computed from the published articles to estimate the reliability of published research results. In a recent study (Bittner and Schönbrodt, 2017), the authors compared the *p*-values and statistical power in two psychology journals with very different IFs (0.79 vs. 5.03) for signs of *p*-hacking and other questionable research practices. In this study, a selection of three different metrics all indicate that the journal with the higher IF published less reliable results than the journal with the lower IF.

## Reproducibility Efforts

Current reproducibility efforts are comparatively small scale, with regard to the number of journals covered in each initiative, and hence cannot provide conclusive evidence as to any differences between journals in the reproducibility of the studies they publish. However, in studies where several journals were covered, the highest ranking journals did not stand out with a particularly high reproducibility, suggesting it may be only average in these journals (Scott et al., 2008; Prinz et al., 2011; Begley and Ellis, 2012).

## CONCLUSION

There are currently several lines of evidence in the literature, suggesting that highly prestigious journals fail to reach a particularly high level of reliability. On the contrary, some of the data seem to indicate that, on average, the highest ranking journals often struggle to raise above the average reliability levels set by the other journals (Table 1).



**TABLE 1** | Overview of the cited literature on journal rank and methodological soundness.

Field	Criteria	Outcome	References
Biomedicine	Image duplications	Higher ranking journals show a lower incidence of image duplications	Bik et al. (2016)
Crystallography	Quality of computer models	Five high-ranking journals significantly below average quality	Brown and Ramaswamy (2007)
Molecular psychiatry	Sample sizes and effect sizes	Higher ranking journals overestimated effect sizes with smaller sample sizes	Munafò et al. (2009)
Neuroscience, psychology	Statistical power	Either no correlation of journal rank with statistical power or a negative correlation	Brembs et al. (2013) Szucs and Ioannidis (2017)
<i>In vivo</i> animal experimentation in disease models	Reporting of randomization and blinded assessment of outcome	Lower reporting of randomization in higher ranking journals and no correlation with reporting of blinded assessment of outcome	Macleod et al. (2015)
Genomics, cognitive neuroscience and psychology	Gene name and <i>p</i> -value errors	More errors in higher ranking journals	Ziemann et al. (2016) Szucs and Ioannidis (2016)
Medicine	Criteria for evidence-based medicine	Two studies found that higher-ranking journals met more criteria, while two failed to detect such an effect	Obremskey et al. (2005) and Lau and Samman (2007) Bain and Myles (2005) and Tressoldi et al. (2013)
Psychology	Three reliability metrics: P-Curve, TIVA and R-index	All three metrics indicate that the higher ranking of two journals publishes less reliable work	Bittner and Schönbrodt (2017)
Biomedicine	Reproducibility of experiments	Reproducibility is low, not even "top" journals stand out	Scott et al. (2008), Prinz et al. (2011) and Begley and Ellis (2012)

Applying criteria such as peer-review, methodology and spread of journal rank covered, the studies cited in the top six rows would be considered as better supported than the bottom three rows.

In particular, comparing higher with lower ranked journals, two main conclusions can be drawn: (1) experiments reported in high-ranking journals are no more methodologically sound than those published in other journals; and (2) experiments reported in high-ranking journals are often less methodologically sound than those published in other journals.

Interestingly, not a single study provides evidence for the third option of higher-ranking journals publishing the **most** sound experiments. It is this third option that one would expect at least one area of inquiry to have conclusively demonstrated, if there was a true positive association between journal rank and reliability.

At the time of this writing, there is one single publication reporting that image duplication is lower in higher ranking journals (Bik et al., 2016). This result conflicts with the increased rate of fraud reported for higher ranking journals (Cokol et al., 2007; Fang and Casadevall, 2011; Fang et al., 2012). Potentially, these contradictory results may indicate that higher ranking journals may be more effective specifically in detecting duplicated images (perhaps due to a superior/more expensive software solution?), but failing in virtually all other aspects.

Importantly, these conclusions have been drawn from evidence collected from the published, unretracted literature, sampling many thousands of publications and data-sets from a variety of experimental fields. This method excludes differences in readership and associated scrutiny and directly approaches a reliability-based notion of methodological "quality".

In total, none of the reported effects of journal rank, even those nominally significant, indicate a clear, obvious difference between journals. All journals seem to have difficulties coping with the issue of reliability, regardless of perceived prestige,

regardless of the selectivity of the journal. Thus, the most conservative interpretation of the available data is that the reliability of scientific results does not depend on the venue where the results was published. In other words, the prestige, which allows high ranking journals to select from a large pool of submitted manuscripts, does not provide these journals with an advantage in terms of reliability. If anything, it may sometimes become a liability for them, as in the studies where a negative correlation was found. This insight entails that even under the most conservative interpretation of the data, the most prestigious journals, i.e., those who command the largest audience and attention, at best excel at presenting results that *appear* groundbreaking on the surface. Which of those results will end up actually becoming groundbreaking or transformative, rather than flukes or frauds, is a question largely orthogonal to the journal hierarchy.

It is up to the scientific community to decide if the signal-to-noise ratio in these journals is high enough to justify the cost of serial scandals and, in the case of medical journals, loss of life, due to unreliable research.

This body of evidence complements evolutionary models suggesting that using productivity as selection pressure in hiring, promotion and funding decisions leads to an increased frequency of questionable research practices and false positive results (Higginson and Munafò, 2016; Smaldino and McElreath, 2016). Arguably, scientists who become "successful" scientists by increasing their productivity through reduced sample sizes (i.e., as a consequence, reduced statistical power) and by publishing in journals with a track record of unreliable science, will go on teaching their students how to become successful scientists. Already after only one generation of such selection pressures, we begin to see the effects on the reliability of the

scientific literature. If scholars strive to convince the public that the scientific endeavor deserves its trust and funds, abolishing these selection pressures is likely the most urgent evidence-based policy of any current reform efforts.

## REFERENCES

- Adler, R., Ewing, J., and Taylor, P. (2009). Citation statistics. *Statist. Sci.* 24, 1–14. doi: 10.1214/09-STS285
- Bain, C. R., and Myles, P. S. (2005). Relationship between journal impact factor and levels of evidence in anaesthesia. *Anaesth. Intensive Care* 33, 567–570.
- Begley, C. G., and Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533. doi: 10.1038/483531a
- Bik, E. M., Casadevall, A., and Fang, F. C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *MBio* 7, e00809–e00816. doi: 10.1128/mBio.00809-16
- Bittner, A., and Schönbrodt, F. (2017). *Assessing the Evidential Value of Journals with P-Curve, R-Index and TIVA*. Nicebread. Available online at: <http://www.nicebread.de/p-curving-journals/> [accessed May 9, 2017].
- Brembs, B., Button, K., and Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Front. Hum. Neurosci.* 7:291. doi: 10.3389/fnhum.2013.00291
- Brown, E. N., and Ramaswamy, S. (2007). Quality of protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 63, 941–950. doi: 10.1107/S0907444907033847
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Cokol, M., Iossifov, I., Rodriguez-Esteban, R., and Rzhetsky, A. (2007). How many scientific papers should be retracted? *EMBO Rep.* 8, 422–423. doi: 10.1038/sj.embor.7400970
- Fang, F. C., and Casadevall, A. (2011). Retracted science and the retraction index. *Infect. Immun.* 79, 3855–3859. doi: 10.1128/IAI.05661-11
- Fang, F. C., Steen, R. G., and Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proc Natl. Acad. Sci. U S A* 109, 17028–17033. doi: 10.1073/pnas.1212247109
- Gordon, M. D. (1982). Citation ranking versus subjective evaluation in the determination of journal hierarchies in the social sciences. *J. Am. Soc. Inf. Sci.* 33, 55–57. doi: 10.1002/asi.4630330109
- Hernán, M. A. (2009). Impact factor: a call to reason. *Epidemiology* 20, 317–318; discussion 319–320. doi: 10.1097/EDE.0b013e31819ed4a6
- Higginson, A. D., and Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biol.* 14:e2000995. doi: 10.1371/journal.pbio.2000995
- Lau, S. L., and Samman, N. (2007). Levels of evidence and journal impact factor in oral and maxillofacial surgery. *Int. J. Oral Maxillofac. Surg.* 36, 1–5. doi: 10.1016/j.ijom.2006.10.008
- Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., et al. (2015). Correction: risk of bias in reports of *in vivo* research: a focus for improvement. *PLoS Biol.* 13:e1002301. doi: 10.1371/journal.pbio.1002301
- Moed, H. F., and van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature* 381:186. doi: 10.1038/381186a0
- Munafò, M. R., Stothart, G., and Flint, J. (2009). Bias in genetic association studies and impact factor. *Mol. Psychiatry* 14, 119–120. doi: 10.1038/mp.2008.77

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

- Obremskey, W. T., Pappas, N., Attallah-Wasif, E., Tornetta, P. III., and Bhandari, M. (2005). Level of evidence in orthopaedic journals. *J. Bone Joint Surg. Am.* 87, 2632–2638. doi: 10.2106/jbjs.e.00370
- Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10:712. doi: 10.1038/nrd3439-c1
- Rossner, M., Van Epps, H., and Hill, E. (2007). Show me the data. *J. Cell Biol.* 179, 1091–1092. doi: 10.1083/jcb.200711140
- Saha, S., Saint, S., and Christakis, D. A. (2003). Impact factor: a valid measure of journal quality? *J. Med. Libr. Assoc.* 91, 42–46.
- Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., et al. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph. Lateral Scler.* 9, 4–15. doi: 10.1080/17482960701856300
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ* 314, 497–497. doi: 10.1136/bmj.314.7079.497
- Smaldino, P. E., and McElreath, R. (2016). The natural selection of bad science. *R. Soc. Open Sci.* 3:160384. doi: 10.1098/rsos.160384
- Sønderstrup-Andersen, E. M., and Sønderstrup-Andersen, H. H. K. (2008). An investigation into diabetes researcher's perceptions of the journal impact factor—reconsidering evaluating research. *Scientometrics* 76, 391–406. doi: 10.1007/s11192-007-1924-4
- Szucs, D., and Ioannidis, J. P. A. (2016). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *bioRxiv* doi: 10.1101/071530
- Szucs, D., and Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biol.* 15:e2000797. doi: 10.1371/journal.pbio.2000797
- Tressoldi, P. E., Giofré, D., Sella, F., and Cumming, G. (2013). High impact = high statistical standards? Not necessarily so. *PLoS One* 8:e56180. doi: 10.1371/journal.pone.0056180
- Vanclay, J. K. (2011). Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics* 92, 211–238. doi: 10.1007/s11192-011-0561-0
- Yue, W., Wilson, C. S., and Boller, F. (2007). Peer assessment of journal quality in clinical neurology. *J. Med. Libr. Assoc.* 95, 70–76.
- Ziemann, M., Eren, Y., and El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. *Genome Biol.* 17:177. doi: 10.1186/s13059-016-1044-7

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Brembs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.