

*Review***A journal veracity–diligence index**

Bor Luen Tang✉

Department of Biochemistry, Yong Loo Lin School of Medicine, National University Health System, National University of Singapore, Singapore

bchtbl@nus.edu.sgorcid.org/0000-0002-1925-636X

Received: 9 Feb 2024
Revised I: 2 Apr 2024
Revised II: 18 Apr 2024
Accepted: 20 Apr 2024
Published: 3 Jun 2024

Conflict of Interest

The author has no conflicts of interest relevant to this article.

Acknowledgements

The author is grateful to all reviewers and editors of *European Science Editing*, whose insightful and constructive comments helped to improve the manuscript.

Funding

This article received no funding.



This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

Citation

Tang BL. A journal veracity–diligence index. *Eur Sci Ed.* 2024;50:e120611.

<https://doi.org/10.3897/ese.2024.e120611>

Abstract

Inaccuracies, false information, and fraudulent work in scientific publications could cause indirect harm, lead to significant negative socioeconomic impacts, and erode public trust in science. Journals – and publishers – play an essential role as gatekeepers in ensuring the veracity of published scientific literature. However, beyond corporate pride and integrity, there is usually no legal obligation or formal regulatory requirement for journals to ensure the veracity of the work they publish or be efficient and transparent in any investigative proceedings. Here, I propose a numerical indicator of the performance of a journal in terms of its efforts at establishing the veracity of the work it publishes and due diligence, an index computed from the following values: a) frequency of alleged irregularities or misconduct, b) frequency of retractions, c) efficiency of the journal's response to concerns or allegations, and d) transparency and thoroughness with which the journal investigates those concerns and announces its findings and actions.

Keywords:

Academic publishing, due diligence by academic journals, research misconduct, retraction of papers, veracity of published research, whistleblowing in publishing

Introduction

Journal and their publishers are key gatekeepers of the quality and veracity of published scientific work.¹ At the pre-publication stage, this function entails proper editorial work and careful engagement of appropriate peer reviewers as well as dedicated in-house efforts to detect dishonest acts such as plagiarism and inappropriate manipulations of data or images in submitted manuscripts. At the post-publication stage, efficient and transparent handling of allegations and reports of irregularities in data or images and other forms of misconduct in published papers and timely issue of notices, corrections, or retractions where appropriate, are significant contributors to the ethos and integrity of science.

However, beyond corporate pride and academic integrity, there is apparently no material incentive for journals to ensure the veracity of the work they publish, nor is there a mandate to be efficient and transparent about investigative proceedings or timely issue of notices of concern or retraction. In fact, given the lack of resources to verify empirical findings in the sciences, editors and reviewers could not be expected to go beyond performing due diligence in scrutinizing submitted data and text.² Such lack of resources also limits effective investigations of allegations of irregularities and misconduct in published papers, and journals often need to involve the authors' institutions,³ leading to delays and opaqueness in the transactions. There is also the question of commercial interest and considerations related to profit. Although a journal wants to publish good papers to enhance its reputation, it is arguable whether there is a corresponding keenness to address, investigate, and remove erroneous or fraudulent papers. Here, I first look briefly at some of these issues and propose a way to measure the efforts at ensuring the veracity

of the work reported in manuscripts and diligence exercised by a journal.

Difficulties in ensuring veracity and in responding to allegations

Holbeach and colleagues have recently expounded on the 'legal obligations that might exist for journal editors and publishing companies with respect to ensuring the veracity of the published literature', and the authors lament that in this regard, 'Ultimately, there is no enforceable legal obligation in Australia, the United Kingdom, or the United States'.⁴ One might add that the same would likely be true for most, if not all, other countries (or the publishing houses would have moved elsewhere). There are also concerns that legalizing such controls may lead to criminalization of research misconduct,^{5,6} the desirability of which is an issue that is debated elsewhere.⁷ In any case, implementation of laws or formulation of mandates to compel journals and publishers to ensure the veracity of what they publish or to be efficient and transparent in investigative proceedings and issue of notices is procedurally cumbersome and fraught with political difficulties.

As far as pre-publication review is concerned, any reputable journal would proclaim that it has stringent peer-review procedures in place. However, many widely publicized cases of research misconduct have shown that fraudulent papers have slipped through the filters of even the most reputable and high-ranking journals.^{8,9} On the other hand, post-publication issues such as concerns and allegations of misconduct would affect all journals big and small. The Committee on Publication Ethics (COPE) does have clear guidelines on how member journals of COPE should respond to concerns raised directly with the journal¹⁰ or through social media.¹¹ However, not all journals are members of COPE, and it is unclear if failure to strictly comply with these

guidelines or failure to do things in a timely manner would have substantially impactful negative consequences for a member journal.

There are expressed grievances in the literature about what was perceived to be inadequate or inefficient response on part of journals and publishers, particularly about issues arising post publication. Analyses have shown that retraction is often slow, and some papers, despite being shown to be fraudulent, remain unretracted for a long period.^{12,13} For example, two papers by Paolo Macchiarini, the disgraced tracheal transplant surgeon, were retracted by the *Lancet* only very recently, more than 5 years after ‘indisputable evidence of falsification in the 2008 paper’.^{14,15} Scientific sleuths¹⁶ have found it difficult to secure retraction of papers that show clear irregularities and misconduct.¹⁷ On occasions, even authors themselves have found it difficult to retract their own papers.¹⁸ Although there are valid reasons for such apparent tardiness – for journals, retractions is a serious matter that must be handled with extreme care – other oversights are not so easily explained.¹⁹

In the light of possible apathy and tardiness associated with the less glamorous part of scholarly publishing and the lack of formal legislative or regulatory measure against them, I hereby propose a simple measure of a journal’s efforts at ensuring the veracity of the work it publishes and of its diligence that offers a semi-quantitative assessment of the degree of stringency in checking manuscripts as well as of the efficiency and transparency of the journal’s responses to concerns related to papers published by that journal.

The journal veracity–diligence index

The journal veracity–diligence index (P_{VD}) is dependent on two opposing sets of parameters, namely (1) the frequency of irregularity or misconduct and of concerns or allegations

raised (F_{im}) and the retraction rate (F_r) and (2) a combination of journal response scores based on efficiency (R_e) and transparency or thoroughness (R_{tt}), and is expressed as their ratio:

$$P_{VD} = (F_{im} + F_r) / (R_e + R_{tt})$$

The term F_{im} is calculated based on the number of papers with verified irregularity or misconduct allegations (but excluding requests for corrections or corrigenda that do not affect the integrity of data or results) for every 1000 papers published by that journal in the past 5 years (per mille, or ‰). The allegations of irregularity or misconduct made against a paper would be largely based on publicly visible posting(s), such as those made on PubPeer.²⁰ Journals should of course declare eventually, after investigation, any such allegations made in confidence. The value of F_r could be obtained from the Retraction Watch Database (RWD),²¹ and the number of papers published can be ascertained from the journal’s website, also expressed per mille. An estimated rate of publications retracted over time made in 2018 is 4 in 10,000, or 0.4‰,²² and this would likely be a lower-boundary estimate of the numbers indicated by $F_{im} + F_r$.

The term R_e is based on a 3-point Likert scale score of how quickly and efficiently a journal acknowledges and responds to, in the form of an editorial note or expression of concern or retraction, allegations of irregularity or misconduct in a published paper (1, no response; 2, slow response; 3, fast and wholesome response). Although an official response from a journal would likely take time for the processing of complaints, investigations, editorial committee deliberations on decisions, and other administrative delays, it would be reasonable to expect that the response be published within a year (or possibly two years under extenuating circumstances). The term R_{tt} is based on a 3-point Likert scale score of the transparency of investigations and details

of the faults or issues as indicated in the retraction notice (and any interim investigation report): 1, generically and minimally worded notice with few details; 2, notice with some description of the data issues or other issues; 3, notice with adequate details of the issues as well as an outline of investigative procedures. Data for R_{it} can be obtained from a close scrutiny of a journal's published notices. Data for R_c would require a more elaborate correlation of the dates when verified concerns or allegations first surfaced and when a response was published by the journal.

A hypothetical example

Take a hypothetical example of a journal that has published 4000 papers over the past 5 years and, over the same period, received verified allegations of irregularity or misconduct for 10 papers, following which 5 were retracted. Therefore, the sum of F_{im} and F_r would be 3.75, and the sum of $R_c + R_{it}$ ranges from 2 to 6. Accordingly, P_{VD} would range from 0.63 to 1.88. A small P_{VD} value indicates that the journal either experiences only a few problems related to the veracity of the work it publishes or is efficient, transparent, and diligent in responding to allegations or both. Conversely, a large P_{VD} value indicates a need for improvement in the above aspects. Using the hypothetical example above, in which the values stipulated for F_{im} and F_r are somewhat conservative, and taking the median value of $R_c + R_{it}$ as 4, the value of P_{VD} will be 0.938.

A real-life example

Here is a real-life example, which analyses the journal *Nature*. During 2019–2023, the journal published 5336 research articles (data from Scopus: <https://www.scopus.com/sources>) and retracted 14 papers (data from Retraction Watch: <http://retractiondatabase.org>). Of these 14 papers, 11 have at least one relevant posting on PubPeer, a number that provided the basal estimate for F_{im} . These retractions

were made fairly promptly (within 1–2 years of publication), and the retraction notices were usually reasonably detailed in outlining the issues that had rendered the papers or their main conclusions untrustworthy. The respective values obtained for each of the parameters (F_{im} , 2.06; F_r , 2.6; R_c , 3; and R_{it} , 3) gave a P_{VD} of 0.78 for *Nature*.

Although a similar analysis of different journals is beyond the scope of the present paper, we compared the results of our analysis with those for a low-impact-factor journal (which shall remain nameless). That journal published 2541 papers during the same period and retracted 46 of them (of which 39 had at least one relevant posting on PubPeer). The retractions were made within 1–2 years of publication, but the retraction notices were generic and seldom provided any details on the reasons for retraction. Besides, many papers are labelled 'withdrawn', but the reasons for the withdrawal are not clearly stated. A conservative estimate of P_{VD} for this journal, based only on RWD records (F_{im} , 15.35; F_r , 18.10; R_c , 3; and R_{it} , 1), is 8.36. This large P_{VD} value is primarily due to the large number of recent retractions, some of which are likely to be due to paper mills being the source of the papers, and is compounded by the lack of detail in the retraction notices.

Discussion

Journal veracity–diligence index as indicator of a journal's diligence before and after publication

The proposed measure (P_{VD} index) reflects a balance between negative consequences of concerns, allegations, and retractions and the timeliness and thoroughness of a journal's responses. The sum of F_{im} and F_r could be zero, because a journal may receive no complaints or allegations and may not have retracted a single paper over a five-year

period. In that case, P_{VD} will also be zero. In reality, this would be rare, except perhaps for predatory journals,^{23,24} which carry out no quality checks. Of course, the same result may also be indicative of a journal particularly stringent in its pre-publication peer review and editing so that none of the papers it publishes ever attracts any concerns or allegations after publication.

However, in academic publishing, oversights and mistakes on part of authors, reviewers, and editors, whether inadvertent or intentional, are not unlikely. For journals specializing in biomedical sciences, or multidisciplinary journals with a good proportion of papers on biomedical sciences, there is a heightened degree of concerns and allegations of misconduct discussed in online forums,^{20,25} and papers *are* retracted, although at varying rates.^{26,27} It would thus be prudent to restrict F_{im} to concerns and allegations of irregularities in data or results and misconduct and to exclude those presenting differing interpretive opinions or critiques or shortcomings of a published paper. Honest mistakes and flaws that are significant or substantial would result in retractions, and thus reflected in F_r .

If occasional oversights and mistakes are inevitable, a journal's response to them assumes importance, and its timeliness is reflected in the terms R_c and R_{it} . It should be clear that P_{VD} is a semi-quantitative index: although F_{im} and F_r are both quantitative terms, R_c and R_{it} , given their nature, are necessarily dependent on more qualitative assessments. The latter could be advantageous in assessing journal qualities that are more complex and difficult to quantify reliably or robustly. Data for R_c and R_{it} could also be reader or third party driven and assessed with data originating, but ultimately evaluated independently, from the journal. With the values for the terms F_{im} , F_r , R_c , and R_{it} in place, aspects of journal quality

associated with efforts at ensuring the veracity of contents, before as well as after publication, should become apparent. These terms would arguably provide added value in assessing a journal's gatekeeper function in scholarly work, going beyond citation counts.

What would be the expected quality-indicating cutoff or optimal range for P_{VD} ? Without specifying any cutoff, a P_{VD} range of 0–1.00 would be reflective of reasonably high standards of veracity and diligence of a journal (but see also the caveats discussed below). Most good and reputable journals should be able to consistently keep P_{VD} within this low range. However, it is envisaged that serious oversights, resulting for example in mass retractions associated with compromised peer review of special issues,^{28,29} would significantly inflate the values of F_{im} and F_r with a drastic rise in P_{VD} . However, barring any recurrence, this unfortunate black mark would be effectively erased after 5 years.

Caveats and limitations

A major criticism against the usefulness of P_{VD} would be the difficulties in implementation. Whereas F_r and R_{it} are more easily obtained, F_{im} is not easy to estimate and might be limited to posts that appear in public forums. Although access to PubPeer records is free, sifting through them would be tedious. This would be greatly facilitated by PubPeer's journal dashboard (with built-in analytical tools) – but that requires a subscription. Hopefully, one could expect journals (or their publishers) that are truly diligent and care about post-publication peer review of their papers to subscribe to the dashboard(s). As to R_c , it would require close scrutiny and correlation of dates and is also highly variable and could differ significantly from paper to paper, because the respective authors should be given, and are usually given, adequate time to defend themselves or to provide evidence to refute the grounds for concerns or allegations.

However, journals could help themselves and the academic community by making the timeline of events transparent, for example by stating clearly when the concerns or allegations first surfaced, the deadlines given to authors to respond, and whether the responses were deemed adequate (resulting in no further action) or inadequate (resulting in an editorial note or expression of concern). Such transparency, as has been proposed for retraction notices,^{30,31} would be helpful and welcomed.

Small F_{im} and F_r values might not necessarily be indicative of stringent pre-publication peer review and editing. It is possible for less well-known journals with limited readership and public attention to receive few or no allegations of misconduct and, therefore, no retractions, despite hosting papers that are erroneous or need to be retracted. This might result in a misleadingly low P_{VD} value compared to that for more prestigious or high-profile journals with wide readership and more frequent post-publication peer review. Furthermore, a journal with high levels of diligence and transparency in dealing with allegations is likely to publish more corrections and retractions than another journal that is less efficient as a gatekeeper. Therefore, a very low P_{VD} score because of smaller values of F_{im} and F_r might not be truly reflective of the journal's efforts to check the veracity of what it publishes; conversely, a journal diligent in gatekeeping might also obtain a high P_{VD} score (despite the latter having a high $R_c + R_{tr}$). It is therefore important not to take the P_{VD} at face value: its component contributing factors must be appropriately scrutinized for any assessment. This limitation of the P_{VD} is perhaps analogous to that of a journal's impact factor, which, for a small journal, can be skewed by one or two papers with inordinately high citations.³² Despite its shortcomings, the P_{VD} as an index is likely to be helpful in gauging the balance between

efforts at checking the veracity before publication and diligence after publication.

Conclusion

Given a lack thereof, and difficulties in achieving, legislative and regulatory measures that would be globally adopted and consensually followed by all journals and publishers, it would probably be more pragmatic to have some deterrent to tardiness while promoting efforts to check the veracity of what is considered for publication (pre-publication stage) and also of what is published (post-publication stage). I have therefore proposed the P_{VD} index as a numerical indicator of efforts at checking veracity and of diligence. I recognize that the semi-quantitative nature of P_{VD} and the difficulties in obtaining robust values for some of its contributing components would impede its wide adoption and usefulness. However, the time has come for the true quality of an academic journal to be assessed using measures other than citation counts, and the P_{VD} index might serve as a modest contribution to this much-needed departure from the conventional approach to assessing the quality of a journal.

If implemented, the P_{VD} can be used not only as a stand-alone indicator of a journal's efforts at checking the veracity of what it intends to publish and of its diligence in monitoring responses to what it has already published but might also be potentially used for imposing a proportionate deduction from other quantitative measures of a journal's standing, such as the Journal Impact Factor (JIF) assigned by Clarivate Analytics or the CiteScore (CS) assigned by Elsevier. Such deduction, for example, may be in terms of a weighted percentage reduction corresponding to the value of P_{VD} . Although much maligned,^{32,33} these metrics continue to be highly sought after by journals and publishers, and any attempt at their devaluation would attract serious

considerations by them. Of course, any formal modification of the JIF or the CS by the P_{VD} would require appropriate interventions from Clarivate analytics and Elsevier, but that should not prevent readers from coming up with some modifications of their own.

References

- Gough N. W(h)ither Gatekeeping? Academic publishing and peer review in complex networked systems. *Access Contemp Issues Educ*. 2012;31(1):29-35.
- Wager E. How should journal editors respond to cases of suspected misconduct? *J Microbiol Biol Educ*. 2014;15(2):146-150. [\[CrossRef\]](#)
- Wager E, Kleinert S, CLUE Working Group. Cooperation & Liaison between Universities & Editors (CLUE): recommendations on best practice. *Res Integr Peer Rev*. 2021;6(1):6. [\[CrossRef\]](#)
- Holbeach N, Freckelton AO QC I, Mol BW. Journal editors and publishers' legal obligations with respect to medical research misconduct. *Res Ethics*. 2023;19(2):107-120. [\[CrossRef\]](#)
- Sovacool BK. Using criminalization and due process to reduce scientific misconduct. *Am J Bioeth*. 2005;5(5):W1-W7. [\[CrossRef\]](#)
- Bülow W, Helgesson G. Criminalization of scientific misconduct. *Med Health Care Philos*. 2019;22(2):245-252. [\[CrossRef\]](#)
- Dal-Ré R, Bouter LM, Cuijpers P, Gluud C, Holm S. Should research misconduct be criminalized? *Res Ethics*. 2020;16(1-2):1-12. [\[CrossRef\]](#)
- Armond ACV, Gordijn B, Lewis J, et al. A scoping review of the literature featuring research ethics and research integrity cases. *BMC Med Ethics*. 2021;22(1):50. [\[CrossRef\]](#)
- Xie Y, Wang K, Kong Y. Prevalence of research misconduct and questionable research practices: A systematic review and meta-analysis. *Sci Eng Ethics*. 2021;27(4):41. [\[CrossRef\]](#)
- COPE. Responding to whistleblowers when concerns are raised directly. Available at: <https://publicationethics.org/resources/flowcharts/responding-whistleblowers-when-concerns-are-raised-directly> Accessed March 2023.
- COPE. Responding to whistleblowers when concerns are raised via social media. Available at: <https://publicationethics.org/resources/flowcharts/responding-whistleblowers-when-concerns-are-raised-via-social-media> Accessed March 2023.
- Trikalinos NA, Evangelou E, Ioannidis JPA. Falsified papers in high-impact journals were slow to retract and indistinguishable from nonfraudulent papers. *J Clin Epidemiol*. 2008;61(5):464-470. [\[CrossRef\]](#)
- McHugh UM, Yentis SM. An analysis of retractions of papers authored by Scott Reuben, Joachim Boldt and Yoshitaka Fujii. *Anaesthesia*. 2019;74(1):17-21. [\[CrossRef\]](#)
- Torjesen I. Lancet will not retract discredited paper on tissue engineered trachea transplants. *Br Med J*. 2022;376:o600. [\[CrossRef\]](#)
- Wilmshurst P. The Lancet's expression of concern – too little, too late. Available at: <https://drpeterwilmshurst.wordpress.com/2023/03/01/the-lancets-expression-of-concern-too-little-too-late/comment-page-1/> Accessed March 2023.
- Oransky I. Meet the scientific sleuths: more than two dozen who've had an impact on the scientific literature. Available at: <https://retractionwatch.com/2018/06/17/meet-the-scientific-sleuths-ten-who-ve-had-an-impact-on-the-scientific-literature/> Accessed March 2023.
- Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. *mBio*. 2016;7(3):e00809-e00816. [\[CrossRef\]](#)
- Sohn R. Two years ago, an author asked a journal to withdraw a paper. It still hasn't. Available at: <https://retractionwatch.com/2023/03/03/two-years-ago-an-author-asked-a-journal-to-withdraw-a-paper-it-still-hasnt/> Accessed March 2023.
- Loadsman JA. Why does retraction take so much longer than publication? *Anaesthesia*. 2019;74(1):3-5. [\[CrossRef\]](#)
- Ortega JL. Classification and analysis of PubPeer comments: how a web journal club is used. *J Assoc Inf Sci Technol*. 2022;73(5):655-670. [\[CrossRef\]](#)
- The Center for Scientific Integrity. The retraction watch database [Internet]. Available at: <http://>

retractiondatabase.org/RetractionSearch.aspx?

Accessed March 2023

22. Brainard J, You J. What a massive database of retracted papers reveals about science publishing's 'death penalty'. Available at: <https://www.sciencemag.org/news/2018/10/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty> Accessed March 2023.

23. Cobey KD, Lalu MM, Skidmore B, Ahmadzai N, Grudniewicz A, Moher D. What is a predatory journal? A scoping review. *F1000Research* 2018;7:1001.

F1000Res. 2018;7:1001. [CrossRef]

24. Grudniewicz A, Moher D, Cobey KD, et al.

Predatory journals: no definition, no defence.

Nature. 2019;576(7786):210-212. [CrossRef]

25. Markie M. Post-publication peer review, in all its guises, is here to stay. *Insights.* 2015;28(2):107-110.

[CrossRef]

26. Yeo-Teh NSL, Tang BL. An alarming retraction rate for scientific publications on coronavirus disease 2019 (COVID-19). *Acc Res.* 2021;28(1):47-53. [CrossRef]

27. Yeo-Teh NSL, Tang BL. Sustained rise in retractions in the life sciences literature during the

pandemic years 2020 and 2021. *Publications.*

2022;10(3). [CrossRef]

28. Stigbrand T. Retraction Note to multiple articles in Tumor Biology. *Tumour Biol.* 2017. [CrossRef]

29. Oransky I. Springer Nature slaps more than 400 papers with expressions of concern all at once. Available at: <https://retractionwatch.com/2021/09/28/springer-nature-slaps-more-than-400-papers-with-expressions-of-concern-all-at-once/> Accessed March 2023.

30. Vuong QH. Reform retractions to make them more transparent. Available at: <https://www-nature-com.libproxyl.nus.edu.sg/articles/d41586-020-01694-x> Accessed March 2023. *Nature.*

2020;582(7811):149-149. [CrossRef]

31. Teixeira da Silva JA, Vuong QH. Fortification of retraction notices to improve their transparency and usefulness. *Learn Publ.* 2022;35(2):292-299. [CrossRef]

32. Dimitrov JD, Kaveri SV, Bayry J. Metrics: Journal's impact factor skewed by a single paper [journal]. *Nature.* 2010;466(7303):179-179. [CrossRef]

33. Callaway E. Beat it, impact factor! Publishing elite turns against controversial metric. *Nature.* 2016;535(7611):210-211. [CrossRef]

ease / publications

ese / European Science Editing

European Science Editing is an official publication of EASE. It is an open access peer-reviewed journal that publishes original research, review and commentary on all aspects of scientific, scholarly editing and publishing.

<https://ese.arphahub.com/>

<https://www.ease.org.uk>

https://twitter.com/Eur_Sci_Ed

<https://www.linkedin.com/company/easeeditors/>



© 2024 the authors. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.