Citation for accepted version: Lee, Carole J., and David Moher. "Promote scientific integrity via journal peer review data." *Science* 357, no. 6348 (2017): 256-257.

Applying the "Trust, but Verify" Model to Journal Peer Review

As the publication landscape shifts, it may be time to apply the "trust, but verify" model to journal peer review.

By Carole J. Lee[1*] and David Moher[2,3]

There is an increasing push by journals to ensure that data and products related to published papers are shared as part of a cultural move to promote transparency and trust in the scientific literature. Yet, few journals commit to evaluating their effectiveness at implementing their own reporting standards (1, 2). Similarly, though the vast majority of journals endorse peer review as an approach to ensure trust in the literature, few of them make their peer review data available to evaluate the effectiveness of the practice towards achieving concrete measures of quality, including consistency and completeness in reporting. Remedying these apparent disconnects is critical for closing the gap between guidance recommendations and actual reporting behavior. We see this as a collective action problem whose redress requires leadership and investment by publishers, where publishers can be incentivized to make these investments through mechanisms that allow them to manage reputational risk and through continued innovation in journal assessment.

[1] Department of Philosophy, University of Washington, Seattle, WA 98195, USA.
* Email: c3@uw.edu.
[2] Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, The Ottawa Hospital–General Campus, Ontario K1H 8L6, Canada.
[3] Visiting Scholar, METRICS, Stanford University, Palo Alto, CA 94304, USA.

The call to open the black box of peer review is decades long, and many concerns raised decades ago (3) still resonate: there is too little sound research on journal peer review; this creates a paradox whereby science journals do not apply the rigorous standards they employ in the evaluation of manuscripts to their own peer review practices; as such, a sound research program on journal peer review is needed to improve the efficiency and effectiveness of scientific communication. Despite longstanding efforts by the International Congress on Peer Review and Scientific Publication to foster and showcase such research, it remains difficult to find experimental studies supporting the huge fiscal and personal investment spent on peer review (4-6).

Large, systematic experimental studies are needed to evaluate how peer review practices can be optimized to achieve transparency and completeness in reporting – especially for those features introduced to enable others to "trust, but verify" that study outcomes have not been selectively introduced, switched, or omitted (e.g., registered reports, open data). At minimum, we need tests for the effectiveness of: different methods for implementing reporting standards into web-based peer review response templates; different methods for training authors, reviewers, and editors; the addition of technologies designed to detect shortcomings in statistical and methodological reporting (7, 8); the number and expertise of peer reviewers involved (e.g., methodologists versus experimentalists in biomedicine). Such research would require a wealth of journal data (e.g., reviewer scores and commentary, manuscript content) and meta-data (e.g., reviewer discipline/expertise) – information too scarcely studied in the growing meta-research literature, which evaluates scientific practice with an eye towards improving it (9).

To address the evidence gap in meta-research on journal peer review, we must recognize that the availability of journal peer review data has been constrained by a collective action problem. Science would be better off if journals allowed for and participated in the empirical study and quality assurance of their peer review processes. However, doing so is resource-intensive and comes at considerable risk for individual journals in the form of unfavorable evidence and bad press. Incentives for journals can run against successful implementation of initiatives designed to prevent spinning the results and discussion of a paper based on excavating to find a significant P-value, as papers reporting statistically significant results tend to garner more citations towards a journal's impact factor (10). Along these lines, editors and publishers report disregarding their own trial registration requirements for fear of losing 'exciting' pap papers to rival journals (11).

How can editors secure resources to evaluate peer review practices at their journals? Publishers can take an active role. They support the efficiency and integrity of journal peer review by providing staff as well as technological infrastructure such as web-based peer review systems and plagiarism-detecting software (e.g., CrossCheck). Publishers can integrate into their peer review systems ways of measuring or auditing the degree to which accepted and rejected peer-reviewed papers meet a journal's reporting standards. More ambitiously, publishers could collaborate with meta-researchers and software designers to create experimental designs and workflows enabling editors/journals to compare competing peer review models/processes. Because journals in the natural, social, and medical sciences have become increasingly consolidated into fewer publishing houses with higher and higher profit margins (12), large publishers are comfortably placed to take the lead on creating, licensing, or

purchasing such systems. Publishers, who develop contracts to distribute author work in accordance with copyright laws, are also well placed to develop agreements that would license publishers and/or meta-researchers to use reviewer scores and commentary towards programmatic evaluation of the journal's peer review practices (13). In the same spirit as the Human Genome Project, which invested 1% of its budget in the Ethical, Legal and Social Implications research program with enormous payoff, we propose an investment on a similar relative scale to enhance the effectiveness of journal peer review.

To incentivize journals to self-assess their peer review practices, it is imperative that they have opportunities to manage the reputational risk associated with conducting and reporting meta-research. One way to do this would be to broaden the content of reporting guidelines to articulate explicit policy expectations for journal meta-research – and do so in ways that give journals the flexibility to make their own choices about how to balance reputational risk against transparency. For example, the second generation of the Transparency and Openness Promotion (TOP) Guidelines could add a new category indicating a journal's willingness to facilitate meta-research on the effectiveness of its own peer review practices. To parallel the three-tier structure of the current TOP Guidelines, this meta-research category could be implemented at different levels of stringency. At the lower end of the spectrum, the journal could disclose whether they are conducting internal evaluations, where journals maintain the ability to retain study results for internal use. In the mid-range, journals could be required to disclose the results of their internal evaluations, where journals could maintain flexibility about how they present their results for external use (e.g., reporting aggregated results across similar journals to pool risk). At the higher end of the spectrum, journals could

relinquish data and analyses to researchers outside the institution for third-party verification – an option that might appeal especially to publishers with fewer resources, as it places the primary financial burden on those conducting the meta-research; alternatively, journals could preregister their experiments in an open access repository and, once research has been completed, deposit their dataset and analyses for public use and verification. This level would be the most open and transparent because it prevents a different kind of selective reporting problem: namely, the failure to report results not favorable to a journal's peer review practices. Adopting any of these levels would signal a journal's commitment to self-assessment and improvement – expectations set by the Committee on Publication Ethics Code of Conduct and Best Practice Guidelines for Journal Editors (14) – as well as a commitment to reducing research waste and maximizing the value of published research (15).

What incentives might journals and publishers have for adopting such a framework? Recent changes in publishing, meta-research, and technology have created new challenges and opportunities. With the rise of illegitimate publishers marketing their journals as "peer reviewed," legitimate journals and publishers must find ways to signal the authenticity and effectiveness of their peer review processes. As the call to de-emphasize journal impact factor gains momentum (16), alternative indicators have been emerging for evaluating journals along different dimensions of intellectual credibility: for example, by the reproducibility (17) or sample size and statistical power of published results (18). By undertaking systematic meta-research, publishers and journals would have the opportunity to certify the legitimacy of their peer review practices – as well as identify and advertise ways that they meet or exceed current reporting practices. Such meta-research can help create a new economy of credibility for

journals, where it is possible for high impact journals to score less well on other measures of intellectual credibility (19). Until such efforts come to fruition, meta-researchers can further incentivize publishers by shifting their focus to compare reporting practices not by journal (20), but by publishing house; and, reviewers can decline to review for journals that neglect to assess their own peer review practices.

Publishers have begun investing in technologies designed to interface with journal data and meta-data to provide review metrics (e.g., number/role of reviewers) and content (e.g., review commentary) that certify the quality and legitimacy of article-level peer review – see, for example, the American Association for the Advancement of Science's (AAAS) recent acquisition of PRE (AAAS is the publisher of Science). PRE – and the trend towards making peer review material available alongside published papers (e.g., EMBO) – do a great service by certifying peer review for individual papers. However, they overlook questions facing the broader scientific community about which peer review practices optimize consistency and completeness in reporting. The goal of the framework we propose is to do just this. Engaging in this work requires workflow integrations and meta-research that have visibility into both accepted and rejected papers. Until then, current practices will unnecessarily enable the publication and citation of inadequately reported studies, entrust the integrity of the scientific corpus to a system of post-publication correction that likely focuses disproportionately on the most visible papers (19), and leave authors in the cross-hairs of highly publicized meta-research studies designed to detect reporting inconsistencies and omissions in published papers. Within the scholarly ecosystem, large publishers are well positioned to undertake such work, not only because of their access to relevant data, but because they are flush at a time when grant

agencies and authors are asked to do more with less. Moreover, cultivating a "reputation of high scholarly values" is a central part of the profitable business of capturing readers and authors (21) – a business sustained by reviewers' good-faith, mostly donated efforts.

As pressures on journals, publishers, and authors continue to shift, it may be time to apply the "trust, but verify" model to journal peer review. Doing so is essential for closing the gap between guidance recommendations and observed reporting behavior. Until then, inadequately reported research will continue to waste time and resources invested by authors, reviewers, journals, academic institutions, funders, study participants, and readers (15) – and limit the credibility and integrity of science.

REFERENCES AND NOTES:

1. A. Rowhani-Farid, A. Barnett, *BMJ Open* **6**: e011784 (2016).
2. N. Vasilevsky, J. Minnier, M. Haendel, M. Haendel, R. Champieux, *PeerJ Preprints* (available at: https://doi.org/10.7287/peerj.preprints.2588v1.
3. J. C. Bailar, K. Patterson, *The New England Journal of Medicine* **312**, 654-657 (1985).
4. R. Bruce, A. Chauvin, L. Trinquart, P. Ravaud, I. Boutron, *BMC Medicine* **14**, 1-16 (2016).
5. S. van Rooyen, F. Godlee, S. Evans, N. Black, R. Smith, *The British Medical Journal* **318**, 23-27 (1999).
6. A. C. Justice, M. K. Cho, M. A. Winker, J. A. Berlin, D. Rennie, *Journal of the American Medical Association* **280**, 240-242 (1998).
7. D. Shanahan, "A peerless review? Automating methodological and statistical review" (http://blogs.biomedcentral.com/bmcblog/2016/05/23/peerless-review-automating-methodological-statistical-review/).
8. N. Nuijten, C. H. Hartgerink, M. A. van Assen, S. Epskamp, J. M. Wicherts, *Behavior Research Methods* **48**, 1205-1226 (2016).
9. J. P. Ioannidis, D. Fanelli, D. D. Dunne, S. N. Goodman, *PLOS Biology* **13**: e1002264 (2015).
10. A.-S. Jannot, T. Agoritsas, A. Gayet-Ageron, T. V. Perneger, *Journal of Clinical Epidemiology* **66**, 296-301 (2013).
11. E. Wager, P. Williams, *BMJ: British Medical Journal* **347** (2013).
12. V. Larivière, S. Haustein, P. Mongeon, *PLOS ONE* **10**, e0127502 (2015).
13. T. Hoke, E. Moyland, on behalf of COPE Council, "Who 'owns' peer reviews? COPE discussion document" (Committee on Publication Ethics, 2016).
14. Committee on Publication Ethics, "Code of conduct and best practice guidelines for journal editors" (publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf).
15. S. Kleinert S, R. Horton, *Lancet* **383** (2014).
16. San Francisco Declaration on Research Assessment (www.ascb.org/dora/).
17. U. Schimmack, "Replicability Ranking of 26 Psychology Journals" (replicationindex.wordpress.com/2015/08/13/replicability-ranking-of-26-psychology-journals/).
18. R. C. Fraley, S. Vazire, *PLOS ONE* **9**, e109019 (2014).
19. F. C. Fang, A. Casadevall, *Infection and Immunity* **79**, 3855-3859 (2011).
20. *The COMPare Trials Project*, B. Goldacre, H. Drysdale, A. Powell-Smith, et al. www. COMPare-trials.org (2016).
21. E. Wager, "Publishing ethics and integrity" in *Academic and Professional Publishing*, R. Campbell, E. Pentz, I. Borthwick, Eds. (Chandos, 2012), chap. 14.

Acknowledgments: Many thanks to M. Baker, C. Bauman, M. McNutt, D. Mellor, B. Nosek, D. Rennie, and the reviewers for comments. CL wrote the initial draft; CL and DM revised. CL and DM are not speaking in their roles as affiliates of these organizations: CL and DM are members of the TOP Coordinating Committee; CL is a contractor for the National Institutes of Health; DM is the Director of the Canadian Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network; and, DM is academic advisor for StatReviewer.