

## 2.6 Relationship between Peer Review and Bibliometrics

Michael Ochsner

**Abstract:** Peer review as the cornerstone of academic quality control has been accused of several biases. As a less subjective and biased alternative, bibliometric methods have been developed and implemented in evaluation procedures. This chapter discusses the relationship between peer review and bibliometrics, showing that bibliometrics is dependent on peer review and not free of its biases. Using three examples of interrelating peer review and bibliometrics, it concludes that rather than playing the two methods of evaluation off against each other, efforts should focus on the interplay and combination of the two.

**Keywords:** Open Science, altmetrics, informed peer review, bias, citation practices, performance-based research funding, predictive validity, interrater reliability, research quality.

### Introduction

Peer review is generally seen as the cornerstone of academic quality control and is often said to date back to the dawn of modern science in the eighteenth century (e.g. Bornmann, 2011). However, Kronick (1990) argues that peer review, depending on how one defines it, has occurred “ever since people began to identify and communicate what they thought was new knowledge. That is because peer review [...] is an essential and integral part of the process of consensus building” (Kronick, 1990, p. 1321). Whenever the exact start of peer review, peer review is the predominant method of evaluation for allocation of grants, the selection of manuscripts, academic recruitment, institutional evaluation, prizes, study programmes etc. (Bornmann, 2011; Daniel, Mittag and Bornmann, 2007; Ochsner et al., 2020). In their roles of “gatekeepers” (Lamont, 2009) or “guardians of science” (Daniel, 1993), peers thus assure the high standards of scientific endeavour and assign merit.

Being so important in academic life, peer review is criticised regularly and its use is challenged. Often, more objective evaluation methods using bibliometric indicators are suggested because peer review is deemed biased and socially contested and accused of slowing down the publication process (Bornmann and Leydesdorff, 2014). Still, peer review is inherent in all evaluation of scientific work, also in indicator-based approaches, simply because evaluation procedures that do not directly rely on peer review still use data on scientific works that include peer review as a central mechanism of evaluation (Ochsner et al., 2020).

---

**Michael Ochsner**, D-GESS, ETH Zurich, Zurich (Switzerland) and FORS, Lausanne (Switzerland),  
ochsner@gess.ethz.ch

In the following, the relationship between peer review and bibliometrics will be explored, giving a short overview on several issues and challenges peer review is facing and explicating bibliometrics' dependency on peer review. The chapter will show that bibliometrics and peer review should not be seen as mutually exclusive alternatives but as complementary methods of evaluation.

## Biases associated with Peer Review

While peer review has been the most established method of evaluation of scientific works at least since the eighteenth century, it has nevertheless been criticised strongly and has come under pressure, especially during the last thirty years, given the availability of biblio- and scientometric data (Bornmann and Leydesdorff, 2014). At the same time, the criticism of peer review is not unchallenged. Rather, peer review is still seen as the most adequate form of evaluation of scientific merit (Hicks, 2012). In the following the literature on biases in peer review is summarised, with the arguments then presented for peer review.

### Different Types of Biases

In the literature, several types of criticisms of the peer review process are discussed (Bornmann, 2011; Lee et al., 2013): The first, *low interrater reliability*, refers to when reviewers disagree on whether a manuscript is good enough for publishing or a project merits funding. The low agreement between reviewers is seen as a problem of the reliability of the process (Bornmann and Daniel, 2008). Second, *issues of fairness* of peer review are raised: criteria other than the merit of the manuscript, project or research might influence the decision, such as gender, the prestige of the author's or applicant's institution, their country or their (former) supervisor or the language of the article, meaning that native English speakers are rejected less often (Cronin, 2009; Wennerås and Wold, 1997). Third, *low predictive validity*, claims that reviewers' judgements might not be linked to the work's later appreciation, for example if reviews for highly cited papers were not more positive than those for less cited ones (Gottfredson, 1978) or if there are no differences between funded or rejected grant proposers regarding citation success (Melin and Danell, 2006). Fourth, the *efficiency* of peer review is questioned as it takes time to review and revise. This can delay research, is said to inhibit innovation and puts much burden on the scientific community (Cowen et al., 1987; Eysenck and Eysenck, 1992). Fifth, the *conservative bias* suggests that peer review leads to preference for established knowledge as peers tend to prefer research similar to their own and experts are often older than submitting authors (Lee et al., 2013). Sixth, experts might prefer research from their own discipline and are often not cognisant in related fields, which might punish research that crosses disciplinary boundaries, leading to *de-valuing of interdisciplinary research*

(Langfeldt, 2006). Seventh, the review process can be *stressful and frustrating*, especially for new authors, and thus can keep talents from doing research (Eysenck and Eysenck, 1992). Women are thereby more vulnerable to this process and seem to let their work be more influenced by evaluators, potentially damaging their own profile as a researcher (Lendák-Kabók and Ochsner, 2020). Eighth, review is said to lead to demonstration of *positive outcome bias*, meaning that only if an outcome is found is it published, while if an expected outcome is not found it is not published, leading to biased reporting, or so-called publication bias (Lee et al., 2013).

### Issues with the Biases

While criticism is an integral part of scientific knowledge production and helps improve the peer review process, many scholars challenge the biases often attributed to peer review. Regarding the often-cited issue of fairness of peer review, especially regarding discrimination of women, many studies cannot replicate such biases (Friesen, 1998; Mutz, Bornmann and Daniel, 2012). Also, at least for prestigious journals, manuscripts rejected but published elsewhere receive less citations than those accepted (Bornmann et al., 2011), contesting the claim of low predictive validity. Langfeldt et al. (2015), however, criticise that studies on biases in peer review, especially regarding fairness and predictive validity, compare outcomes without having a clear concept of what the outcome should be: for example, does one really expect that a person who receives an open-mode research grant will be relatively more cited after the grant? Isn't it likely that the person is already chosen because of a higher impact potential? Ochsner (2020) goes further and challenges the methodology behind the studies. First, a high interrater reliability is not necessarily desirable as it might just be a sign of an unfortunate choice of experts following the same paradigm and thus rejecting research drawing from another paradigm. Without the reasons behind the different ratings, it cannot be interpreted as bias. Second, a high predictive validity, usually measured by citations, might just point to the fact that being published in this specific journal or having received a prestigious grant bolsters the citation rate of the article or scholar. Citations are not a good outcome indicator and not a valid measure for a functioning peer review process because citing a source can have many different meanings (Tahamtan and Bornmann, 2018). Third, if in a peer review process a bias can be identified, it might not be the problem inherent in the peer review but might lie in conditions external to it: e.g. if women are less self-confident and submit understated proposals or researchers at prestigious institutions have more time to write proposals and men work more often at prestigious institutions, peer review would be in favour of men in both conditions even though the peer reviewers would not favour men as such (see, e.g., Ceci and Williams, 2011; Enserink, 2015). Lipworth et al. (2011) even argue that the social and subjective dimensions of peer review are in fact the very essence of peer review, simply because the decisions to be taken are always choices against the background of many valua-

ble, sometimes contradictory, information and as gatekeepers, reviewers and editors are expected to play this social role (see also Eysenck and Eysenck, 1992).

Most interestingly, even the authors cited above evoking the biases of the review process and also the researchers surveyed or interviewed in several studies on the perceptions of the peer review process confirm that peer review is still seen as the best, or least bad, way of improving research (be it manuscripts or research proposals) or gatekeeping (Cowen et al., 1987; Eysenck and Eysenck, 1992; Lendák-Kabók and Ochsner, 2020; Vanholsbeeck, 2020).

## Links between Bibliometrics and Peer Review

Given that bibliometric assessments are advertised as a less biased and costly alternative to peer review, especially among policymakers (Taylor, 2011), it is worthwhile to reflect on the links between peer review and bibliometrics. First, some critics of peer review base their argument upon bibliometric measures: the amount of citations or correlations with citations are used to “validate” peer review procedures. However, citations have not been validated as good measures of scientific performance (Ochsner, Hug and Daniel, 2012) and it seems tautological to argue that bibliometrics are better suited because peer review outcomes do not correspond to bibliometric outcomes. Second, it is obvious that there is a strong link between peer review and bibliometrics because the main data sources for bibliometric analysis, Web of Science and Scopus, include only peer-reviewed publications. Additionally, performance-based research funding models relying on comprehensive national publication databases include only publications that were peer-reviewed (Verleysen and Engels, 2013). Third, indicator-driven performance-based research funding models can also involve a peer review component: peers decide which publication channels are considered first or second level (Sivertsen, 2016). Thus, bibliometric measures are not free of the biases ascribed to peer review.

Therefore, the question should not be whether bibliometrics should replace peer review but how to disburden and improve peer review processes, and evaluation procedures in general; and bibliometrics can play a role in this quest. Peer review can take many roles in evaluation procedures, appearing in different forms, and the decisions taken by the peers can be of varying significance (Ochsner 2020). Therefore, the interlinks between bibliometrics and peer review should come more into focus. I mention three examples worthy of further investigation.

### Open Peer Review

With Open Science as a main policy goal, new versions of peer reviewing and publishing attract interest. Open Access (OA) journals that are not published in print anymore are not limited in space. Reviews can be made public and linked to different

versions of the same article. Different models of Open Peer Review processes have been identified (Ross-Hellauer, 2017). This leads to new possibilities for bibliometric indicators, taking into account hitherto mostly hidden aspects of scientific work: the writing of reviews. Publons, recently integrated into Clarivate's Web of Science, already implemented a new indicator, the ratio of articles published and reviews submitted (publons.com). Other indicators are likely to emerge as well. However, it is most likely they will suffer from similar problems as other bibliometric measures, such as issues of coverage, disciplinary differences and the risk of gaming.

### **Ambiguity of Peer-reviewed Publications**

Given that peer review is an important cornerstone of academic quality assurance, it also takes on an important role in evaluation practices. Whether a publication is peer-reviewed is often a criterion in evaluation procedures, for example performance-based funding systems, recruitment or grant funding; researchers are thus often asked to identify their peer-reviewed publications. However, even experts disagree whether a certain journal or book publisher applies peer review (Pölönen, Engels and Guns, 2020). This has consequences for the use of peer-reviewed publications in performance-based evaluation systems, which is why a label for peer-reviewed publications was created (Verleysen and Engles, 2013). It is somewhat amusing that an indicator for peer review helps assure the quality of a bibliometric indicator.

### **Informed Peer Review**

However, not only peer review can be used to improve or innovate bibliometric indicators. Vice versa, bibliometric and scientometric indicators can also be used to improve peer review. As research quality is a complex construct, reviewers often disagree in their judgement, not because they disagree about single aspects of quality but because they apply different weightings of those aspects (Eysenck and Eysenck, 1992). Moreover, given the often high volume of works or profiles to evaluate, indicators assigned to aspects of research quality can help to take an informed decision. Ochsner, Hug and Daniel (2014) therefore suggest assigning indicators to different aspects of research quality and letting reviewers rate each aspect. Such a procedure will lead to more reliable and fair judgements (Thorngate, Dawes and Foddy, 2009).

## **Conclusions**

The evaluation of research is a complex endeavour. Neither bibliometrics nor peer review are without problems or flaws. Instead of playing peer review and bibliomet-

rics off against each other, more efforts should be spent on how the two interact and how they are best combined.

## References

- Bornmann, L & Leydesdorff, L 2014, 'Scientometrics in a changing research landscape', *EMBO reports*, vol. 15, no. 12, pp. 1228–32.
- Bornmann, L, Mutz, R, Marx, W, Schier, H & Daniel, H-D 2011, 'A multilevel modelling approach to investigating the predictive validity of editorial decisions: do the editors of a high profile journal select manuscripts that are highly cited after publication?', *Journal Of The Royal Statistical Society Series A-Statistics In Society*, vol. 174, no. 4, pp. 857–79.
- Bornmann, L 2011, 'Scientific Peer Review', *Annual Review of Information Science and Technology*, vol. 45, no. 1, pp. 199–245.
- Bornmann, L & Daniel, H-D 2008, 'The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie*', *Angewandte Chemie International Edition*, vol. 47, no. 38, pp. 7173–78.
- Ceci, SJ & Williams, WM 2011, 'Understanding current causes of women's underrepresentation in science', *Proceedings of the National Academy of Sciences*, vol. 108, no. 8, pp. 3157–62.
- Cowen, EL, Spinelli, A, Hightower, AD & Lotyczewski, BS 1987, 'Author reactions to the manuscript revision process', *American Psychologist*, vol. 42, no. 4, pp. 403–05.
- Cronin, B 2009, 'Vernacular and vehicular language', *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, pp. 433.
- Daniel, H-D 1993, *Guardians of science: Fairness and reliability of peer review*, VCH Verlagsgesellschaft.
- Daniel, H-D, Mittag, S & Bornmann, L 2007, 'The potential and problems of peer evaluation in higher education and research', in A Cavalli (ed), *Quality assessment for higher education in Europe*, pp. 71–82, Portland Press, London.
- Enserink, M 2015, 'Dutch sexism study comes under fire', *Science*, vol. 364, no. 6446.
- Eysenck, HJ & Eysenck, SBG 1992, 'Peer review: Advice to referees and contributors', *Personality and Individual Differences*, vol. 13, no. 4, pp. 393–99.
- Friesen, HG 1998, 'Equal opportunities in Canada', *Nature*, vol. 391, no. 6665, p. 326.
- Gottfredson, SD 1978, 'Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments.', *American Psychologist*, vol. 33, no. 10, pp. 920–34.
- Kronick, DA 1990, 'Peer review in 18th-century scientific journalism', *Journal of the American Medical Association*, vol. 263, no. 10, pp. 1321–22.
- Lamont, M 2009, *How professors think: Inside the curious world of academic judgment*, Harvard University Press.
- Langfeldt, L, Bloch, CW & Sivertsen, G 2015, 'Options and limitations in measuring the impact of research grants – evidence from Denmark and Norway', *Research Evaluation*, vol. 24, no. 3, pp. 256–70.
- Langfeldt, L 2006, 'The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments', *Research Evaluation*, vol. 15, no. 1, pp. 31–41.
- Lee, CJ, Sugimoto, CR, Zhang, G & Cronin, B 2013, 'Bias in peer review', *Journal of the American Society for Information Science and Technology*, vol. 64, no. 1, pp. 2–17.
- Lendák-Kabók, K & Ochsner, M 2020, 'A gender and geopolitical perspective on peer review', in M Ochsner, N Kancewicz-Hoffman, M Hołowiecki & J Holm (eds.), *Overview of peer review practices in the SSH. ENRESSH Report. European Network of Research Evaluation in the Social Sciences and Humanities*, pp. 78–86.

- Lipworth, WL, Kerridge, IH, Carter, SM & Little, M 2011, 'Journal peer review in context: A qualitative study of the social and subjective dimensions of manuscript review in biomedical publishing', *Social Science & Medicine*, vol. 72, no. 7, pp. 1056–63.
- Melin, G & Danell, R 2006, 'The top eight percent: development of approved and rejected applicants for a prestigious grant in Sweden', *Science and Public Policy*, vol. 33, no. 10, pp. 702–12.
- Mutz, R, Bornmann, L & Daniel, H-D 2012, 'Does Gender Matter in Grant Peer Review?', *Zeitschrift für Psychologie*, vol. 220, no. 2, pp. 121–29.
- Ochsner, M 2020, 'Place, role, form and significance of peer review in national research evaluation systems', in M Ochsner, N Kancewicz-Hoffman, M Hołowiecki & J Holm (eds.), *Overview of peer review practices in the SSH. ENRESSH Report. European Network of Research Evaluation in the Social Sciences and Humanities*, pp. 55–60.
- Ochsner, M, Hug, SE & Daniel, H-D 2014, 'Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies', *Zeitschrift für Erziehungswissenschaft*, vol. 17, no. 6, pp. 111–32.
- Ochsner, M, Hug, SE & Daniel, H-D 2012, 'Indicators for Research Quality in the Humanities: Opportunities and Limitations', *Bibliometrie – Praxis und Forschung*, vol. 1, p. 4.
- Ochsner, M, Kancewicz-Hoffman, N, Holm, J & Hołowiecki, M 2020, 'Conclusions', in M Ochsner, N Kancewicz-Hoffman, M Hołowiecki & J Holm (eds.), *Overview of peer review practices in the SSH. ENRESSH Report. European Network of Research Evaluation in the Social Sciences and Humanities*, pp. 99–100.
- Pölönen, J, Engels, TCE & Guns, R 2020, 'Ambiguity in identification of peer-reviewed publications in the Finnish and Flemish performance-based research funding systems', *Science and Public Policy*, vol. 47 no. 1, pp. 1–15.
- Ross-Hellauer, T 2017, 'What is open peer review? A systematic review', *F1000Research*, vol. 6, no. 588.
- Sivertsen, G 2016, 'Publication-Based Funding: The Norwegian Model', in M Ochsner, SE Hug & H-D Daniel (eds.), *Research Assessment in the Humanities. Towards Criteria and Procedures*, pp. 79–90, Springer International Publishing, Cham.
- Tahamtan, I & Bornmann, L 2018, 'Core elements in the process of citing publications: Conceptual overview of the literature', *Journal of Informetrics*, vol. 12, no. 1, pp. 203–16.
- Taylor, J 2011, 'The Assessment of Research Quality in UK Universities: Peer Review or Metrics?', *British Journal of Management*, vol. 22, no. 2, pp. 202–17.
- Thorngate, W, Dawes, RM & Foddy, M 2009, *Judging merit*, Psychology Press, New York, NY.
- Vanholsbeek, M 2020, 'Peer review in the context of the new modes of knowledge production, dissemination and evaluation', in M Ochsner, N Kancewicz-Hoffman, M Hołowiecki & J Holm (eds.), *Overview of peer review practices in the SSH. ENRESSH Report. European Network of Research Evaluation in the Social Sciences and Humanities*, pp. 87–93.
- Verleysen, FT & Engels, TCE 2013, 'A label for peer-reviewed books', *Journal of the American Society for Information Science and Technology*, vol. 64, no. 2, pp. 428–30.
- Wennerås, C & Wold, A 1997, 'Nepotism and sexism in peer-review', *Nature*, vol. 387, pp. 341–43.

