

Evaluation of the Quality of Scientific Research

Vlatko Silobrčić

Donald School Journal of Ultrasound in Obstetrics and Gynecology (2019): 10.5005/jp-journals-10009-1602

INTRODUCTION

Let me begin by trying to define what is to be evaluated. The task is to evaluate quality of scientific research, be it research proposals and/or research results. The quality, in this context, depends primarily on the creativity of scientists and engineers, in the activity often defined by the popular syntagma of "Research and Development" (fully: "Scientific Research and Experimental Development"). Although the syntagma includes both scientists and engineers, I will concentrate on scientific research, both fundamental and applied.

Thus, in essence, I would like to deal with approaches to evaluate creativity of scientists. For all I know, there is no better way to do it but to have relevant experts express their opinion about the quality of the proposal and/or of results. This "peer review" (further PR) is, in my view, inescapable in evaluating not only scientific but also other creative human activities.

With the advance of information technologies, numerical scientometric and bibliometric indicators have been developed and often added to PR. If used with caution and understanding of the meaning of these indicators, they can help to obtain an objective assessment of the quality of scientific texts. Although the numbers are attractive for measuring any activity, I would like to stress that they should be used to complement PR rather than replace it.¹

Continuing, I will try to describe most of the accepted tools presently in use for evaluation of the quality of scientific research. Let me start with PR and add some numerical indicators to supplement it. My goal is to present their principal characteristic and provide some data for the comparison among them. Whenever possible, I will add a relevant literature for further information on the subject.

PEER REVIEW

Peer review is the process of trying to obtain an objective opinion of competent colleagues. Their opinion helps decision makers (e.g., journal editors) to reach a documented decision based on an expert judgment.

It is expected of the reviewers to, first of all, establish the originality (new knowledge) of the proposed project/manuscript. This expert judgment then helps to decide if a proposed text is worth financing/publishing. Sometimes it may be justified to consider as original, and worth publishing, a confirmation of the published original finding. An expert in the field should recognize that instance and clearly argument the need to publish such confirmation.

Next are ethical issues to be considered: has the author(s) acknowledged the contribution of other colleagues working in the same field of research? Also, e.g., in medical research, ethical issues regarding patients have to be taken into account. Similarly, questions of animal protection may be relevant.

Finally, reviewer(s) have to establish if the text is presented clearly and briefly and with proper argumentation. It is expected

Department of Natural Sciences, Croatian Academy of Sciences and Arts, Zagreb, Croatia, European Union

Corresponding Author: Vlatko Silobrčić, Department of Natural Sciences, Croatian Academy of Sciences and Arts, Zagreb, Croatia, European Union, Phone: +385 91 461 43 23, e-mail: vlatko.silobrcc@gmail.com

How to cite this article: Silobrčić V. Evaluation of the Quality of Scientific Research. *Donald School J Ultrasound Obstet Gynecol* 2019;13(4):159–161.

Source of support: Nil

Conflict of interest: None

that the expert indicates to the editors of scientific journals any errors in clarity and brevity and inconsistencies of the sentences, with her/his proposal for improvement.

Thus, peer, in the context of PR, means that the evaluator is competent in the particular field, i.e., well informed of the recent developments in field of research for which she/he is expected to act as a reviewer. The other very important characteristic of the reviewer is objectivity—impartiality. A responsible reviewer will recognize her/his potential conflict of interest and decline to do PR in such instances. It is an unwritten rule that PR is done as a courtesy to academic colleagues, and it is rarely paid for.

One of the problems of the procedure is that it takes some time and it delays the decision on the submitted manuscript and/or project. I can tell from my own experience that it may take more than a day to properly evaluate an average manuscript. Given the flood of information produced by the ever-increasing number of scientists, it is not always easy to find a proper colleague to do the review within a required time, since the good ones are often too busy. The advance of the communication technology is helping a lot to speed the process, to the delight of editors of scientific journals and other media that distribute the new knowledge.

Historically, PR has had a number of proposed variations: from completely abolishing it to variations of anonymity and access to the general public. As the ideal of open access is penetrating the scientific community, there are consistent ideas to adapt the process to openness. The desirable ideal is to do PR anonymously to try to assure objectivity. It is true, however, that the anonymity is not always achievable. It is often possible to discover the authors of the texts evaluated, either true the references cited or other similar ways. However, it seems to me that anonymity should remain a desirable goal in spite of the fact that it could be compromised.

The other extreme is the possibility of publishing the PR assessment together with the published manuscript. And even the discussion between the author(s) and the reviewer. This is definitely in line with the ever-increasing openness of scientific endeavor (we should not forget that the investment in science often comes from public sources). As far as I know, neither of the new models of PR

has been tested properly to be accepted as the ultimate solution by the scientific community. Principles of the “open access” are presented in my review article.²

As far as I know, such proposals have not been generally accepted. I firmly believe that this “filter” before providing financing, particularly with public funds, and enabling publication of research results, is extremely important to assure acceptable quality in science. This is even more important in view of the fact that the number of scientists and results of their activity are constantly increasing, and the concomitant but unwelcome pressure to “publish or perish” often required for the advancement in scientific career.

Unfortunately, in PR involvement of colleagues brings along the danger of interfering human relations between the evaluator and the evaluated. Objectivity in the process, on the other hand, is definitely the essential ingredient of PR. This is particularly critical within small scientific communities. The only true way of avoiding this type of potential conflict of interest is to take entire scientific community as the pool of evaluators. In most cases, this means using the language that functions as the prevailing common language of science—English.

Although the PR is constantly discussed within the scientific community,^{3,4} additional tools for evaluating research have come into use, based on the advancement of information technologies. Let me therefore include some of them under this general title.

SCIENTOMETRIC INDICATORS

The advance of the computer technology has opened the way to apply this possibility to devise numerical indicators for “measuring” scientific research. In my view, these indicators can be used to assist the review process, but it has to be done with extreme caution not to try to replace the PR.¹ In other words, the reviewer has to be aware what any given indicator means to use it properly. For example, is it and to what extent possible to take the “impact factor” of a scientific journal in which a scientific paper is published, as the ultimate measure of the quality of a given text?

Numbers can help in some simple situations. For example, a simple count of published text can function if one has to compare two individuals or institution of which one has zero productivity and the other has any number of published works. But to say that a scientist who has published 10 papers is automatically more creative/successful than one with 5 papers, without any further qualifications, would be a *vitium artis*.

Number of citations has been often used as a substitute for the quality of scientific research. Here again one has to be very careful: citations where, by whom, in positive or negative context, etc. Again, the expert opinion is required to add proper insight into the matter.

Impact factor of the journal in which a scientific paper was published is often taken as an indicator of the quality of the published research results. The initial assumption is that a journal that has a high impact factor is showing the quality of the journal and, by the same talking, assures the high quality of the publication process (competent editorial board, strict selection of submitted papers, etc.). To the certain extent, this holds true. Since I have written a review article on the impact factor, I offer it to colleagues more interested in the details.⁵ It may serve the purpose of this text to repeat the main conclusion of the paper. The impact factor of a given journal, let me remind you, is calculated by dividing the number of published papers in a given period of time by the

number of citations that the journal receives in the next 2 years after the paper was published. There are possibilities for refining the factor by increasing the time of including the accumulated citations, since this makes the average number of citations more correct, with less yearly variations. In whatever way calculated, the impact factor is an average valid for the journal. It does not hold true for every paper published in the journal. It is known that a large proportion of published papers are never cited. So, there may be a paper published in a high-impact-factor journal that is never cited or at least below the impact factor for the journal. It would be erroneous to ascribe the impact factor of the journal to that single paper and its author(s).

But even if the citations to the journal are representing the citations to the paper, one will have to be sure how to ascribe the contribution of each author to the success of the paper. In this, some journals ask of authors to describe singular contribution to the given paper. Is it always the first author the main contributor? The final order of several authors may vary in different fields of science. In many instances the first author is the junior author, while the main author is at the end of the list. Such considerations, of course, are relevant when evaluating the entire activity of a single scientist or a group of them.

h-index

The index has been introduced by Hirsch.⁶ It has been defined as follows: “A scientists has index h if h of his/hers N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each” (as cited by Harzing and van der Wal).⁷ It is a single number and preferable to other single numbers, like number of papers, total number of citations, and citations per paper, since it measures both quantity (number of papers) and quality (number of citations), according to Glaenzel.⁸ Let us take, as an example, that an author has an h -factor of 20 for his published papers. This simply means that, among the published papers, there are 20 of them cited at least 20 times at a given time.

One problem with the h -index is that, if any single paper receives very high number of citations above the index, this will not influence the h -index, since it will be included among others. Therefore, the g -index has been proposed to correct for this possible outlier. Glaenzel⁸ has examined opportunities and limitations of the h -index.

g-index

To deal with this shortcoming of the h -index, the g -index was proposed by Egghe.⁹ In case of the h -index, if a single paper within the included papers receives much higher citations than the h -index, this will not show. The g -index takes this into account and ascribes to such extremes at least “ g ” citations. As far as I know, the g -index has not been generally used, in spite of the fact that it may be a useful correction of the h -index.

Relative Scales

I believe that the concept of relative indicators can be useful for comparing individuals, institutions, and countries by using the same indicator, and finding the position of the evaluated group within a range of given indicator taken for comparison.¹⁰⁻¹² I found this concept very useful in determining the position of an individual/institution/country in comparative environment. For example, when trying to select a standing of an individual or a group for purposes of financing, advancement in scientific carrier, scientific awards, etc.

What one has to do first is have a clear vision of the final goal. For example, if one is to find the best candidate for a scientific award,

it is necessary to define which part of the scientific activity is in question. Let us say that we would like to establish which among the individual scientists deserves an award for her/his achievements. To aid the PR evaluation, we might want to find out how the candidate stands in comparison with another candidate. Finding the relevant indicator and forming a scale of a given group of scientists, the indicator ascribed to a given name can be positioned on the formed scale. This way one can arrive at the relative position of the candidate among her/his competitors. Moreover, together with a proper peer expertise, one can arrive at the relative position of a single candidate. This relative position can provide good argument in favor of the decision of the expert. And, as I have indicated at the beginning, the same process could be used to aid the PR in a number of occasions. I find this approach very useful in conditions where the small scientific communities may have need for proper general criteria for evaluation.

COMPARING THE DATABASES IN ASSISTANCE TO EVALUATE SCIENTISTS AND THEIR RESULTS

One of the comprehensive studies of the comparison has been done by Harzing and van der Wal.⁷ They concluded that the use of Google Scholar (GS) results in “more comprehensive coverage in the area of management and international business” including books, conference papers, and non-USA journals, of which many are not covered by Web of Science (WoS) and Journal Citation Reports (JCR). The WoS and JCR provide data for calculating the Journal Impact Factor (JIF). In their study, the alternative metrics provided by GS showed a strong correlation with the JIF but were more inclusive. The authors also introduced the h-factor and the g-factor to correct for the possibility that a single highly cited paper adds considerably to the mean scores of a journal or an individual scientist. In their view, the h-index for authors testifies more to a “sustained and durable research performance.” The same is true if the h-index is used for journals. On the basis of these results, the authors “strongly suggest” to use the GS impact indices for the “more accurate measure of the true impact.” Also, being freely available it is more accessible to those with less financial means.

INSTEAD OF A CONCLUSION

A number of other authors have contributed to the sensitive field of evaluation of scientists. Thus, Bornmann et al.¹³ explored standards of good practice for analyzing bibliometric data and presenting and interpreting results. Harnad,¹⁴ a renowned researcher in the field, has compared performance metrics with PR. Giske¹⁵ has studied benefits of bibliometry to evaluation of science. Similarly, Leydesdorff¹⁶ described how citation-based journal indicators add to the bibliometric toolbox. Nicolini and Nozza¹⁷ discussed the objective evaluation of scientific performance worldwide. Todd and Ladle¹⁸ examined hidden dangers of the “citation culture.” Zitt and Bassecouard¹⁹ have undertaken to discuss challenges for scientometric indicators.

I do hope that my text is clear enough to convey the message that the field of evaluating scientific research is a complex endeavor

with many aspects and uncertainties. With this in mind, instead of giving a firm conclusion about the practice of evaluating scientific research, I have decided to provide additional literature (see above) to those readers who would like to dig deeper into the field, and wish them all the best in their attempt. I do hope that I have provided principles of theoretical grounds for the evaluation of scientific research. Practice is even more challenging.

REFERENCES

1. Silobrić V. Can bibliometric and scientometric indicators be useful for evaluating the quality of scientific research. *Eur Sci Ed* 2001;27:102.
2. Silobrić V. Open access to peer reviewed scientific texts-A desirable future for informing scientists. *Periodicum biologorum* 2005;107(1):117–121.
3. Peer Review-Its present and future stand, Conference Report, www.esf.org, 2006.
4. Evaluation of Research and Research Funding Organisations. A report by ESF Member Organisation Forum on Evaluation of Publicly Funded Research, www.esf.org, 2012.
5. Silobrić V. How to increase the impact factor of a journal. *Donald Sch J Ultrasound Obstet Gynecol* 2015;9(4):357–360. DOI: 10.5005/jp-journals-10009-1422.
6. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Nat Acad Sci USA* 2005;102(46):16569–16572. DOI: 10.1073/pnas.0507655102.
7. Harzing A-WK, van der Wal R. Google scholar as a new source for citation analysis. *Ethics Sci Env Polit* 2008;8:61–73. DOI: 10.3354/ese00076.
8. Glaenzel W. On the opportunities and limitations of the h-index. *Sci Focus* 2006;1:10–11.
9. Egghe L. Theory and practice of the g-index. *Scientometrics* 2006;69:131–152. DOI: 10.1007/s11192-006-0144-7.
10. Schubert A, Braun T. Relative indicators and relation charts for comparative assessment of publication output and citation impact. *Scientometrics* 1986;9(5–6):281–291. DOI: 10.1007/BF02017249.
11. Schubert A, Glaenzel W, Braun T. Subject field characteristic citation scores and scales for assessing research performance. *Scientometrics* 1987;12(5–6):267–272. DOI: 10.1007/BF02016664.
12. Schubert A. Using the h-index for assessing single publication. *Scientometrics* 2009;78:59. DOI: 10.1007/s11192-008-2208-3.
13. Bornmann L, Mutz R, Neuhaus C, et al. Citation counts for research evaluation: standards of good practice for analysing bibliometric data and presenting and interpreting results. *Ethics Sci Env Polit* 2008;8:93–102. DOI: 10.3354/ese00084.
14. Harnad S. Validating research performance metrics against peer review. *Ethics Sci Env Polit* 2008;8:103–107. DOI: 10.3354/ese00088.
15. Giske J. Benefitting from bibliometry. *Ethics Sci Env Polit* 2008;8:79–81. DOI: 10.3354/ese00075.
16. Leydesdorff L. How are new citation-based journal indicators adding to the bibliometric toolbox? *J Amer Soc Informat Sci & Tech* 2009;60(7):1327–1336.
17. Nicolini C, Nozza F. Objective assessment of scientific performances world-wide. *Scientometrics* 2008;76(3):527–541. DOI: 10.1007/s11192-007-1786-9.
18. Todd PA, Ladle RJ. Hidden dangers of a “citation culture”. *Ethics Sci Env Polit*. 2008;8:13–16. DOI: 10.3354/ese00091.
19. Zitt M, Bassecouard E. Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. *Ethics Sci Env Polit* 2008;8:49–60. DOI: 10.3354/ese00092.