Arch. Immunol. Ther. Exp., 2009, **57**, 13–18
PL ISSN 0004-069X
DOI 10.1007/s00005-009-0001-5

New developments in the use of citation analysis in research evaluation

Henk F. Moed

Center for Science and Technology Studies, Leiden University, the Netherlands

Received: 2008.12.01, **Accepted:** 2008.12.22, **Published online:** 2009.02.14

Abstract

This paper presents an overview of research assessment methodologies developed in the field of evaluative bibliometrics, a subfield of quantitative science and technology studies, aimed to construct indicators of research performance from a quantitative statistical analysis of scientific-scholarly documents. Citation analysis is one of its key methodologies. The paper illustrates the potentialities and limitations of the use of bibliometric indicators in research assessment. It discusses the relationship between metrics and peer review; databases used as sources of bibliometric analysis; the pros and cons of indicators often applied, including journal impact factors, Hirsch indices, and normalized indicators of citation impact; and approaches to the bibliometric measurement of institutional research performance.

Key words: impact factor, Hirsch index, citation analysis.

Corresponding author: Henk F. Moed, Center for Science and Technology Studies, Leiden University, the Netherlands, e-mail: moed@cwts.leidenuniv.nl

INTRODUCTION

In most OECD countries there is increasing emphasis on the effectiveness and efficiency of government-supported research. Governments need systematic evaluations for optimizing their research allocations, reorienting their research support, rationalizing research organizations, restructuring research in particular fields, or augmenting research productivity. In view of this they have stimulated or imposed evaluation activities. Evaluative bibliometrics is a subfield of quantitative science and technology studies aimed at constructing indicators of research performance from a quantitative analysis of scholarly documents. Citation analysis is one of its key methodologies. Citation analysis involves the construction and application of a series of indicators of the "impact", "influence", or "quality" of scholarly work, derived from citation data, i.e. data on references cited in footnotes or bibliographies of scholarly research publications.

This paper focuses on the assessment of the contributions scientists and scholars make in their research publications to the advancement of valid scientific-scholarly knowledge. It recognizes its crucial importance for global economic progress and social welfare, but at the same time it acknowledges that a firm political or societal basis for "basic" research can be maintained only by further developing a system of internal quality control and performance enhancement. Bibliometric indicators provide useful tools in such a system.

The structure of this paper is as follows. Section "General issues: Bibliometric indicators and peer review" deals with general principles underlying bibliometric methodologies for research assessment and discusses the relationship between such methodologies on the one hand and peer review on the other. Section "Databases and their coverage: Web of Science versus Scopus" presents two databases of scientific-scholarly literature that play a key role as data sources in bibliometric research assessment: Thomson-Reuters' Web of Science and Elsevier's Scopus. Section "Indicators" critically discusses three bibliometric indicators that are often used in studies of research performance: the journal impact factor (IF), the Hirsch index (H-index), and a normalized or "relative" citation impact indicator. Finally, section "Concluding remarks" makes general and concluding comments.

GENERAL ISSUES: BIBLIOMETRIC INDICATORS AND PEER REVIEW

The use of citation analysis in the evaluation of individuals, groups, and institutions is more appropriate the more it is:

- formal – i.e. it is previously known to evaluators or decision makers and to scholars or institutions subjected to evaluation that indicators are used as one of the sources of information
- open – those subjected to bibliometric analysis have the opportunity to examine the accuracy of the underlying data and provide background information that in their view is relevant for a proper interpretation of the quantitative outcomes
- scholarly founded – that bibliometric investigators present their outcomes within a scholarly framework, discuss issues of validity, explicitly state theoretical assumptions, and underline their potentialities and limits
- supplemented with expert and background knowledge about the substantive content of the work under evaluation, the conditions under which the evaluated scholars operated, and their research objectives
- carried out in a clear policy context – i.e. applied within the framework of an evaluation procedure of which both the evaluative perspective and the objectives are clear to all participants
- stimulating users to explicitly state basic notions of scholarly quality, its dimensions, and how they were operationalized and weighted
- enlightening rather than formulaic – the indicators are used to obtain insight into a particular aspect addressed in the process rather than as inputs to formulas designed to algorithmically generate the process' outcomes.

In principle it is valid to interpret citations in terms of intellectual influence. Citing authors tend to ensure that important groups and their programs are represented in the reference list of their papers. Including works in a reference list can be interpreted in terms of intellectual influence, even though its expression in the citing text may be vague or implicit. However, it needs to be emphasized that the concepts of citation impact and intellectual influence do not coincide. Outcomes of citation analysis of basic science research groups tend to correlate statistically in a positive way with peer ratings of the groups' past performance. However, the correlations are not perfect.

The future of research assessment endeavors lies in the intelligent combination of metrics (including bibliometric indicators) and peer review. Policy makers, peer review committees, and bibliometric researchers share the responsibility to develop assessment methodologies combining quantitative and qualitative approaches in a proper way. For instance, policy makers may let the type of peer review they plan to carry out in a particular discipline depend upon the outcomes of a bibliometric study of the groups active in that discipline or peer committees may use citation analysis for initial rankings of the groups under evaluation and explicitly justify in their reports why their final judgments deviated from these rankings. In this way, the peer review process may become more transparent.

DATABASES AND THEIR COVERAGE: WEB OF SCIENCE VS SCOPUS

The Citation Indexes produced by Thomson Reuters, especially the Web of Science (WoS) and its predecessor, the Science Citation Index (SCI), are the most frequently used databases in bibliometric studies. In fact, these databases have properties that which make them most appropriate for bibliometric analysis; they have multi-disciplinary coverage, they tend to cover leading international peer-reviewed journals, they have an internal coverage monitor, they process journals cover-to-cover, they include cited references from each source article, and they include all authors of a source article as well as their institutional affiliations. Currently the WoS covers about 9500 journals, but this number is expected to increase in the coming years.

The extent to which citation analysis based on the WoS can be validly applied in all domains of scholarship, including the applied and technical sciences, social sciences, and humanities, is often debated. The WoS does not claim to have complete journal coverage, but rather to include the most important. Their founder, Eugene Garfield, developed a powerful and unique criterion for expanding the database beyond the core of journals whose importance in a given field is obvious: the frequency at which journals are cited in those sources that are already included in the index (Garfield 1964; Garfield 1972; Garfield 1979).

Applying a "database internal" criterion, it can be shown that WoS coverage tends to be excellent in physics, chemistry, molecular biology, biochemistry, biological sciences related to humans, and clinical medicine; good, yet not excellent, in applied and engineering sciences, biological sciences related to animals and plants, geosciences, mathematics, psychology, and other social sciences related to medicine and health; and moderate in other social sciences, including sociology, political science, anthropology, educational sciences, and, particularly, the humanities (Moed 2005).

A principal cause of non-excellent coverage is the importance of sources other than international journals, such as books and conference proceedings. In fields with a moderate coverage, language or national barriers play a much greater role than they do in other domains of science and scholarship. In addition, research activities may be fragmented into distinct schools of thought, each with their own "paradigms".

In 2004, Elsevier launched Scopus, a multidisciplinary citation index covering some 15,000 journals and proceedings volumes. Since its launch, bibliometric researchers have started exploring the potentialities and limitations of this new database for bibliometric analysis. More and more studies are being published comparing the WoS and Scopus. Some of these studies also consider a third citation index launched in the same year as Scopus: Google Scholar.

Any comparison of the WoS and Scopus is hampered by the fact that both are in continuous development. Source coverage is expanded, backlogs are added, and data capturing and standardization are improved. Specific outcomes may therefore become quickly obsolete.

The outcomes of a comparison of WoS and Scopus coverage at the level of individual articles indicates that in science-related fields the overwhelming part of articles and reviews in journals covered by the WoS is also included in Scopus. The overall percentage of WoS-covered science-related papers found in Scopus increased over the years, from 89% in 1996 to 97% in 2005. In other words, in these fields and for published articles and reviews, the WoS constitutes almost a genuine subset of Scopus. On the other hand, Scopus includes about 50% more papers than the WoS (Moed and Visser 2008).

A comparison of WoS and Scopus coverage of the "best" publications submitted to the 2001 Research Assessment Exercise showed that Scopus coverage is especially better in the subject groups Subjects Allied to Health and to a lesser extent also in Engineering and Computer Science and Health Sciences. In Clinical Medicine, Biological Sciences, and Physical Sciences, however, Scopus coverage is slightly lower than WoS coverage. The Scopus database used in these analyses is based on raw data provided by the Scopus team on August 1, 2007. Additions and corrections made after this date are not included. The WoS database used in the analysis is based on annual deliveries of raw WoS data by Thomson Scientific up to and including 2006 (Moed and Visser 2008).

These findings suggest that the criteria for selecting sources are rather different for the two databases (Lopez-Illescas et al. 2009a; Lopez-Illescas et al. 2009b). The WoS's coverage is primarily based on Eugene Garfield's concept of measuring the importance of journals on the basis of their citation impact and including the most important ones as sources in the database. Scopus coverage is more comprehensive and the citation impact of journals is apparently less discriminative, although it includes the overwhelming part of the WoS journals in science-related fields.

More research into the quality of the sources indexed by Scopus across research fields is needed in order to obtain a better understanding of its coverage and its usefulness for evaluative-bibliometric purposes. Nevertheless, even at this stage of development the conclusion seems justified that Scopus is a genuine alternative to the WoS as a data source for bibliometric indicators of research performance in science-related fields.

INDICATORS

In science, the research group is the "natural" unit of scientific activity and constitutes the most appropriate unit of analysis in institutional research performance assessment. Several types of bibliometric indicators are applied to assess the research performance of individual researchers, research groups, and departments. This paper discusses three of these: journal IF, H-index, and normalized citation impact indicators.

The journal IF

The journal IF developed by Eugene Garfield and published by Thomson Reuters in the Journal Citation Reports (JCR) is probably the most widely dispersed bibliometric construct. It was used by Eugene Garfield to select the most important scientific journals for inclusion in the SCI. Their importance is assessed through a combination of an objective and truly unique internal monitor based on citation relationships among journals with assessments by experts from the various fields. One of the indicators applied in the internal monitor is nowadays known as the journal IF.

However, journal IFs are inadequate measures of a group's research performance. first of all, those published in the JCR are not always accurate (Moed and Van Leeuwen 1996), to some extent manipulable (Reedijk and Moed 2008), and strongly affected by differences in citation practices among research fields (e.g. Moed 2005). Even if one applies indicators that correct for such inaccuracies and biases, and even though journal impact is a performance aspect in its own right, journal impact cannot be used to predict the actual citation impact of a group's published articles (e.g. Garfield 1996; Seglen 1994).

The last point is illustrated in Table 1. It shows for a sample of about 2000 authors from the UK the amount of variance in the actual citation impact of an author's papers (using a normalized indicator of citation impact presented below) explained by the average IF of the journals in which the group has published and the average normalized impact of the citation impact of these journals (see below). It shows that the average JCR impact factor explains only 11% of the variation and the normalized journal impact indicator 23%.

Table 1. Correlation between actual citation impact and journal impact for 2000 authors

Indicator	Explained variance in actual citation impact
No. of published articles	0%
Average journal IF	11%
Normalized journal IF	23%

Table 2. H-index for three authors

Author 1		Author 2		Author 3	
no. cites	papers	no. cites	papers	no. cites	papers
30	P1	30	P1	300	P1
10	P2	10	P2	100	P2
8	P3	8	P3	8	P3
6	P4	6	P4	6	P4
5	P5	5	P5	5	P5
1	P6	4	P6	1	P6
0	P7	4	P7	0	P7
		4	P8		
		4	P9		
H=5		H=5		H=5	

The H-index

The H-index, an appealing construct, is highly biased towards "older" researchers with long careers and those active in fields with high citation frequencies and provides an incomplete picture of a group's actual citation impact. There is a huge literature on the mathematical properties and the pros and cons of the H-index.

Table 2 presents the publication lists of three different authors. For each author the papers are ranked by the number of received citations and are symbolized by the character P followed by their rank number. The first author has a publication oeuvre of 7 papers. Five of these are cited at least 5 times. Therefore, the H-index for this author amounts to 5. The second author has 7 papers with citation frequencies equal to those of the first author, but also a large number of papers cited 4 times. There are good grounds for the claim that the performance of the second author is higher than that of the first. However, the H-index for the second author is also 5. The third author has published two papers receiving a large number of citations, much larger than the two top papers of the first or the second author. However, his H-index is 5, identical to that of the other two. This simple case shows that distinct citation distributions can generate the same value of the H-index, while it is questionable whether they reflect the same performance.

A normalized or relative indicator of citation impact

A normalized or relative indicator of citation impact can be defined as the average citation rate of the papers published by a unit of assessment divided by the world citation average in the scientific subfields in which the unit is active. The version of this type of indicator applied in studies carried out at CWTS does not only correct for differences in citation practices among scientific subfields, but also for differences in the expected citation frequencies of the various types of papers (reviews tend to be cited more frequently than articles) and for differences in the "age" of cited papers (older papers can be cited during a longer time period and therefore tend to have higher citation rates than more recently published ones).

Figure 1 presents a cognitive profile of a research group active in the field of medical pharmacology. The horizontal axis indicates the share of its publication output that is published in journals assigned to the various subject categories indicated along the vertical axis. Behind the names of the subject category and in parentheses the normalized citation impact is indicated of the papers published in a particular subject category. For instance, about 37% of the publications made by the research group were published in journals assigned to the subject category Neurosciences. The average citation impact of these papers compared with the world citation average in the subject category, also taking into account the type of paper and their age, amounts to 1.42.

Table 3 illustrates how the normalized citation impact of a research group is calculated. In this example

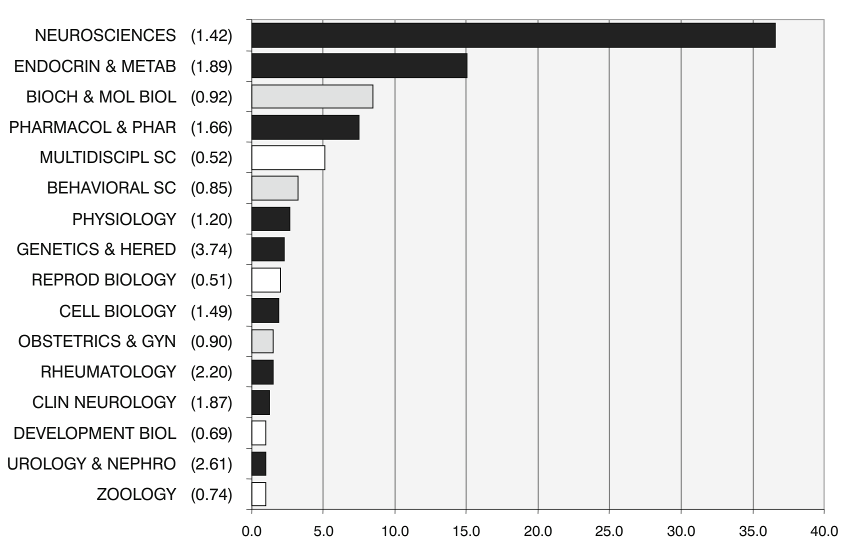

Fig. 1. A cognitive profile of a research group in medical pharmacology.

Table 3. Example of the calculation of a department's normalized citation impact

Document type	Department		Subfield
	papers	citations per paper	citations per paper
Article	10	5.0	4.0
Review	2	10.0	8.0
Normalized citation impact=(10×5.0+2×10.0)/(10×4.0+2×8.0)=70/56=1.25			

a group has published 10 articles and 2 reviews in a particular subfield. Their average citation rates are 5.0 and 10.0, respectively. The total actual number of citations collected by the group's papers amounts to 10×5.0+2×10.0=70. In the entire subfield in which the group is active, articles and reviews are assumed to be cited on average 4.0 and 8.0 times, respectively. The expected number of citations received by the group's papers can be calculated as 10×4.0+2×8.0=56. The normalized citation impact is defined as the ratio of actual and expected number of citations and in this example amounts to 1.25. When a group has published papers in two subfields, a similar type of weighting scheme is applied, the weights being determined by the number of papers in each subfield. For more technical details, the reader is referred to Moed et al 1995, and for more examples to Van Raan 2004. The notion that actual citation rates of groups or journals must in some way be related to the citation characteristics or averages for the fields in which they are active can be found in several publications (e.g. Narin 1976; Vinkler 1986; Braun et al. 1988).

CONCLUDING REMARKS

Bibliometric indicators are useful tools in the assessment of research performance provided that they are accurate, sophisticated, up-to-date, combined with expert knowledge, and interpreted and used with care. During the past two decades, the Leiden Center for Science and Technology Studies has conducted dozens of citation analyses of scientific institutions, research departments, research groups, and individual scientists. The outcomes were used as additional information in evaluating their research performance. These studies were carried out along the following main lines.

- In a first step the time period of analysis was fixed and a list was compiled of the scientists who were active in the entities to be evaluated.
- Their names were matched with a publication database containing all source articles processed for the ISI Citation Indexes, and for each name a preliminary list of publications was compiled.
- The preliminary lists were sent to the scientists involved for verification. Missing articles were added and incorrectly assigned papers were deleted.
- The verified lists were subjected to a citation analysis. Both simple and sophisticated bibliometric indicators were calculated and special analyses were carried out within the framework of the particular policy issues addressed in the study.
- The outcomes of the citation analyses were sent to the scientists subjected to the analysis for comments, enabling them to provide background information that was in their view indispensable for a proper interpretation of the results.
- The bibliometric results, the scientists' comments, and "tentative" conclusions by the analysts were included in a final report. This report was sent to the agency undertaking the evaluation.
- In most studies, a public report was written presenting the main outcomes and conclusions, at a high level of aggregation. Smaller entities subjected to the analysis, such as research departments, were anonymous.

Acknowledgment: This paper is partly based upon my book "Citation Analysis in Research Evaluation", Kluwer, 2005.

REFERENCES

Braun T, Glänzel W, Schubert A (1988) World flash on basic research – the newest version of the facts and figures on publication output and relative citation impact of 100 countries 1981–1985. Scientometrics 13:181–188

Garfield E (1964) "The Citation Index" – A new dimension in indexing. Science 144:649–654

Garfield E (1972) Citation analysis as a tool in journal evaluation. Science 178:471–479

Garfield E (1979) Citation Indexing. Its theory and application in science, technology and humanities. Wiley, New York

Garfield E (1996) How can impact factors be improved? BMJ 313:411–413

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 102:16569–16572

López-Illescas C, De Moya-Anegón F, Moed HF (2009a) Coverage and citation impact of oncological journals in the Web of Science and Scopus. J Informetrics (in press). Available online via: http://dx.doi.org/10.1016/j.joi.2008.08.001

López-Illescas C, De Moya-Anegón F, Moed HF (2009b) Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus: The effect of poorly cited journals in oncology. J Inf Sci (in press).

Moed HF (2005) Citation Analysis in Research Evaluation. Kluwer Academic Publishers, Dordrecht

Moed HF, de Bruin RE, van Leeuwen TN (1995) New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications. Scientometrics 33:381–442

Moed HF, van Leeuwen TN (1996) Impact factors can mislead. Nature 381:186

Moed HF, Visser MS (2008) Appraisal of citation Data sources. A report on a study within the Framework Agreement for the development of a new research assessment and funding system, commissioned by the Higher Education Funding Council for England (HEFCE). Available via: http://www.hefce.ac.uk/Pubs/rdreports/2008/rd17_08/

Narin F (1976) Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. National Science Foundation, Washington D.C.

Reedijk J, Moed HF (2008) Is the impact of journal impact factors decreasing? J. Document 64:183–192

Seglen PO (1994) Causal relationship between article citedness and journal impact. J Am Soc Inf Sci 45:1–11

Van Raan AFJ (2004) Measuring Science. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems. Kluwer Academic Publishers, Dordrecht, pp. 19–50

Vinkler P (1986) Evaluations of some methods for the relative assessment of scientific publications. Scientometrics 10:157–178