Clinics and Research in Hepatology and Gastroenterology (2011) 35, 336—337

Available online at ScienceDirect www.sciencedirect.com
Elsevier Masson France EM|consulte www.em-consulte.com/en

EDITORIAL

Research and bibliometrics: A long history...

The objective evaluation of research activities is not a recent problem. One of the missions of the United States' National Science Foundation (created in 1950) [1,2] was to assess the technological and scientific levels of the major industrialized countries. In 1958, Eugene Garfield founded the Institute for Scientific Information [3] — an organization that subsequently set up the Science Citation Index [4] (SCI, the first ever large database of references) and introduced the Impact Factor [5] (the first ever prestige indicator for scientific journals). The number and complexity of bibliometric and scientometric indicators have continued to grow since then, with a view to providing an automated way to classify countries (the SCImago ranking [6], for example), universities (the Shanghai ranking [7], for example) or researchers in terms of scientific quality and productivity.

In France, the Observatory for Science and Technology [8] has been the country's benchmark organization in this field for 20 years. It has published many reports on France's scientific ranking in the world. Similarly, the major research institutions (Inserm, CNRS, etc.) have long used indicators and dashboards to help them assess their research groups' performance levels.

Basic concepts

Bibliometric indicators are calculated over a period of time (usually between 3 and 5 years) and generally use two inputs:

- the number of publications — an indicator that measures productivity;
- the number of citations — an indicator that measures the impact of the articles produced.

The indicators' reliability is nevertheless conditioned by two major aspects:

- the choice of databases. There are many bibliometric databases; some are multidisciplinary (e.g. SCI and Scopus [9]) and others are specific for a particular area. The number of referenced journals varies significantly from one database to another, as does the percentage of ''national'' journals (French journals, for example) [10];
- identification of publications on the basis of the addresses provided by the authors: this approach is not simple, since addresses are rarely standardized. It is necessary to validate addresses before calculating indicators.

On a macroscopic basis, bibliometric indicators are robust and reliable for comparing or ranking large organisations or geographic areas. However, the problem becomes more complex when assessing individual researchers or research groups.

Individual assessment

Bibliometric assessment has gradually shifted from a macroscopic level to an individual level. However, data validation is essential prior to any calculation of bibliometric indicators. In the case of individual assessment, it is essential to address:

- the problem of homonyms: this issue can partly be resolved by the use of addresses in databases such as Scopus and SCI;
- multiple names: in many countries, women's maiden names and married names have to be taken into account.

The best possible practice would consist in calculating indicators after validation of publication lists by the researchers concerned.

Once the data have been validated, one can calculate indicators. In addition to traditional indicators (the numbers of publications and citations), many other indicators have been suggested for individual assessment. These indicators may involve several complementary concepts:

- position in the author list: in some disciplines (the life sciences, for example), position in the author list is

doi:10.1016/j.clinre.2011.04.008

very important and must be taken it into account when computing indicators;

- the impact factor of the journal in which the publication appeared. We can calculate an impact factor or a total impact factor weighted by the position. Many bibliometric specialists have noted bias in the impact factor [11], including high heterogeneity and widely differing values from one discipline to another. It is also possible to derive indicators that standardize the impact factor by discipline (e.g. the SIGAPS score).

The h-index

The h-index (invented by JE Hirsch [12].) is currently very fashionable indicator. It takes account of both the number of publications and the number of citations per publication: ''*A scientist has index h if h of his/her Np papers have at least h citations each, and the other (Np - h) papers have no more than h citations each*''.

This indicator is used to a variable extent and includes a few sources of bias:

- The h-index depends on the age of the researcher: the older an article, the higher the potential number of citations is likely to be. Hence, a researcher who has a long career is likely to have a higher h-index than a young researcher, even when the latter is more deserving. It is therefore very difficult to compare scientists of different ages. Some authors have proposed to normalize the h-index by the number of years since a researcher's first publication (the m-index);
- Disciplines vary in their citation policy. The number of citations of an article is of value if it is compared with the number of citations for articles in the same discipline (the top 1% or the top 5%, etc.). This is important because the number of citations is the determining factor in calculating the h-index;
- The h-index does not take into account the position in the author list: in some scientific fields (such as physics or mathematics), a publication generally has very few authors and the position in the author list has little importance. This is not the case in medicine, where the position in the list of authors is perfectly codified. In fact, a researcher can achieve a very high h-index without being the first or last author on any of his/her publications; this can make use of the h-index very tricky in medicine.

The many variants of the h-index proposed to date (all based on the number of articles and citations) involve the use of other data to correct a known or suspected source of bias.

Lastly, database choice is also very important. Even though the results may be stable at a macroscopic level, individual-level indicators can differ significantly as a function of the database used. For example, Poynard et al. compared the h-index of 158 hepatologists computed on the basis five different data sources: Google Scholar, Scopus, Web of Science, ScholarL (Scholar limited to ''liver'') and HepaTop (a specialty database). Although there was generally a correct correlation between the five h-indices, the disparities for a given researcher were sometimes large.

Several factors may explain these discrepancies:

- the list of referenced journals: for example, Scopus covers about 18,000 journals, whereas the Web of Science covers nearly 11,000. The two databases share about 8300 journals (i.e. 10,000 are specific to Scopus and 1200 are specific to WoS);
- the list of documents recorded: some databases list letters, editorials and/or conference proceedings as well as full articles;
- the age of the citations: Scopus references citations since 1996, whereas Web of Science goes back much further.

In conclusion, it is clear that bibliometric indicators must be used with caution. It may well be prudent to use several indicators at the same time. This is particularly true in the context of individual assessment, since it seems illusory to summarize a researcher's work by some indicators, however relevant they may be! This was recently emphasized in a report entitled ''The proper use of bibliometrics for individual evaluation of researchers'' that was submitted to the French Minister of Higher Education and Research by the French Academy of Science [13] and which gives a number of basic rules in this area.

References

[1] Godin B. The emergence of S&T indicators: why did governments supplement statistics with indicators. Res Policy 2003:679—91.
[2] http://www.nsf.gov/statistics/.
[3] http://www.isinet.com.
[4] http://thomsonreuters.com/products_services/science/.
[5] Garfield E. The history and meaning of the journal impact factor. J Am Med Assoc 2006;293:90—3.
[6] http://www.scimagojr.com/.
[7] http://www.arwu.org/.
[8] http://www.obs-ost.fr.
[9] http://www.scopus.com/.
[10] Vieira ES, Gomes JANF. A comparison of Scopus and Web of Science for a typical university, Scientometrics; 2009.
[11] Ojasoo T, Maisonneuve H, Matillon Y. Le facteur d'impact des revues, un indicateur bibliométrique à supprimer. Press Med 2002;31:776—82.
[12] Hirsch JE. An index to quantify an individual's scientific research output, Physics; 2005.
[13] Anon C. http://www.academie-sciences.fr/activite/rapport/avis170111.pdf.

P. Devos
EA2694, Department of Research, CHU de Lille, 59000 Lille, France
E-mail address: patrick.devos@univ-lille2.fr.