Metrics in science

Métricas na ciência

We are increasingly surrounded by performance assessment methods in science. These methods, although mathematically correct, are not always able to appropriately reflect reality. These indices assess authors, journals, and classified studies. The Hirsch index, or h index, was developed in 2005 by the physicist J. E. Hirsch at the University of California to assess the cumulative impact and relevance of the research output of an individual[1]; that is, the objective of the h index is to quantify the scientific production of an individual by relevance.

The best way to explain this index is by means of the following example: the h index of the individual will be equal to 10 if the researcher has 10 articles that have been cited at least 10 times each. The effect is logarithmic, that is, to progress from 1 to 2 is much easier than from 11 to 12, because in this case, the number of citations needed will be much higher. There are some limits to this index[2]. The number of citations can increase the h index. This number can be purged in some databases, but if this is done, one loses the power to measure the development of a line of research. Cases were also found of citations deliberately crossed between authors with the aim of leveraging the indices of authors involved, which is not an accurate measurement.

There are other ways to evaluate a particular author with regards to his/her scientific production, such as: evaluating the total number of published articles, evaluating the total number of citations, separately evaluating the number of citations per article or isolating the total number of articles from a large number of citations produced by the author. The h index, nevertheless, represents more globally the relevance of the scientific production of the author.

Another limitation is that this index should be used only to compare authors in the same area or in the same field of science. Specialists in plastic surgery, for example, could be at a disadvantage when compared to cardiologists because cardiology is an area that has a larger amount of indexed journals and, therefore, has a greater probability of being cited in journals without these articles having a higher scientific value. Another potential disadvantage of the h index is that this metric does not give importance to highly cited articles. In fact, the most cited article by Hirsch, in which the author described the h index, had 2196 citations in accordance with the Scopus database, while his h index is "only" 45 (as of March 2015). This demonstrates that the h index is not perfect, but provides a general idea of the productivity of a researcher.

The Impact Factor (IF) of a journal is measured by the ratio between the total number of articles in the journal cited in journals in the last 2 years and the total number of articles published in the journal during the same period. In other words, the greater the citation of articles in relation to the total number of articles, the greater the IF will be. Certainly, an author will prefer to have a study published in a journal with higher visibility, but a journal with higher visibility does not always have the greatest impact. Nevertheless, any author would want an article to have an important impact in medical circles, which can be reflected by the number of citations of a particular article. What makes an article to be more or less cited? This will depend on the subject area in which an article was written, the quality of the article, and where it was published.

Definitely, some journals are not in databases with wider dissemination and, between those of the same database, some journals are more or less influential-this is where the IF counts. A journal will be, theoretically, more influential, to the extent that its articles will have a higher citation in that database. This is theoretical, because the number of citations of an article basically depends more on the article than the journal *per se*[3]. Certainly, reviewers of high quality journals will require better content during the evaluation of an article than reviewers of lower quality journals; therefore, the fact that an article has been accepted by a high quality journal is a good sign, but this is not everything.

DOI: 10.5935/2177-1235.2015RBCP0161

Several defects and vulnerabilities were found in this index. One of them, as previously mentioned, is the fact that a few articles may significantly raise the IF, while others will not even be cited. One way to artificially increase the IF is to publish better quality articles at the beginning of January; this results in generating a high number of citations because the IF begins to be calculated from this month and increases the possibility of citation of an article published in this period. Some editors encourage the creation of supplements on matters whose articles traditionally have a high number of citations. The citation crusade, agreed between publishers of journals, and the fact that some editors encourage authors to cite recent articles published in the same journal are punishable by databases, which has already occurred in the past.

Another way to try to qualify the articles is to classify the types of study according to levels, using the principles of Evidence-Based Medicine (EBM). The articles are classified (in a pyramid) into five types of articles, in accordance with their level of evidence. This classification motivates plastic surgeons to improve the level of scientific production, as the authors themselves should perform a classification of the study during the submission of the article.

As always, there are two sides to a coin. Although this classification is excellent for new articles, when the EBM is used in systematic reviews, success is not always achieved[4]. Surgical areas and interventional medicine are in constant evolution; the assessment of a surgical technique or an interventional technique over a number of years, may not be appropriate, since techniques can be improved and materials refined. All this can lead to a difference in the results over the years. When a meta-analysis is performed on a technique, one can have very well performed studies with level 1 evidence in the 1970s, 1980s, and 2015. At the end there will be a study (meta-analysis) that will be considered as a level 1 evidence, with unreliable results, since the techniques were carried out with a large difference in time and, therefore, with different techniques and materials.

Many other indices seek to evaluate the impact of an article, but these are the ones that have been the most used. Some indexes are attempting to assess articles based on accessions on the web, such as Altimetrics, but can be criticized from a scientific point of view, because internet access is universal and not just restricted to the medical setting.

These metrics deserve a more in-depth analysis, especially with a focus on plastic surgery, in order to provide an idea of what is happening in specialty journals, especially in relation to the quality of the scientific production of other areas. Evaluations are always important guides in giving one an idea of how and where we stand.

REFERENCES

1. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005;102(46):16569-72. DOI: http://dx.doi.org/10.1073/pnas.0507655102
2. Tedesco ACB, Nahas FX, Ferreira LM. The Hirsh (h) Index. Plast Surg Pulse News. 2010;2(3).
3. Rodrigues MA, Tedesco AC, Nahas FX, Ferreira LM. Journal impact factor versus the evidence level of articles published in plastic surgery journals. Plast Reconstr Surg. 2014;133(6):1502-7. DOI: http://dx.doi.org/10.1097/PRS.0000000000000214
4. Nahas FX, Ferreira LM. Evidence-based medicine: the past and the future of plastic surgery. Plast Reconstr Surg. 2014;134(3):499e-500e. DOI: http://dx.doi.org/10.1097/PRS.0000000000000474

FABIO XERFAN NAHAS
Universidade Federal de São Paulo, São Paulo, SP, Brazil
LYDIA MASAKO FERREIRA
Universidade Federal de São Paulo, São Paulo, SP, Brazil