REVIEW ARTICLE JSLS

Scientific Literature and Evaluation Metrics: Impact Factor, Usage Metrics, and Altmetrics

Michael S. Kavic, MD, FACS, Richard M. Satava, MD

ABSTRACT

Evaluating the quality of a scientific article has proven to be an elusive task. The most widely used bibliometric value currently used for this purpose, the journal impact factor, was not originally designed to determine the quality of research in a scientific article. Nevertheless, it has morphed into a surrogate to delineate the quality of a journal and even to represent the quality of individual articles in that that journal. Early 21st century advances in computer technology have seen an explosive revolution in scientific publication that have included open access, online publication, and world-wide accessibility to these publications. These developments have made it obvious that more sophisticated tools are required to delimit the quality of material present in the scientific literature. Usage data, which is measured as the number of full-text downloads of a specific article, is just one new method to evaluate the source of the vast material available that can be leveraged to more fully evaluate the merit of scientific literature.

Key Words: Literature Evaluation Metrics, Impact Factor, Usage Metrics, Altmetrics.

INTRODUCTION

Science has been described as a gift giving economy wherein one's ideas are published and freely shared in order to increase general knowledge and, in some measure, impact the thinking of others. For many years, scientists have used citations to identify source material and indicate which authors influenced their work. Citations also may be used as a metric to capture the impact of a scientific publication.[1]

In the scientific publishing industry, research metrics are tools to measure performance at the journal level and at the author level. No cost to readers open access publications, along with readily available online communication and technology, have revolutionized the accessibility, distribution, and gathering of scientific information. Coupled with this revolution has been an explosion of creative research enabled by these technologies such that clinicians and investigators are challenged to keep up with the huge volume of research activity.[2]

The Journal of the Society of Laparoscopic and Robotic Surgeons (JSLS) – entirely open access without any embargo of scientific material - has experienced a similar explosion involving information distribution. For each of the past six years, JSLS has recorded in excess of one million full-text downloads of its articles each year. The reach and scope of this powerful phenomenon has yet to be fully delineated. But its impact in the distribution of scientific knowledge cannot be dismissed or denied.

As Stewart Brand once famously said "...information wants to be free..."[3] The dispute of whether valuable scientific information should have a cost or should be available for free is not the purpose of this communication. Rather, the movement towards open access of information is acknowledged and, as a consequence of that movement, there has developed a new dimension in the measurements that determine the value of scientific work. Along with these new measurements are different ways to access the citation data and numbers used to calculate a scientific work's value.

CITATION METRIC

Journal Impact Factor

For nearly 50 years, the only tool available to assess journal performance was the citation metric, impact factor (IF), also known as the journal impact factor (JIF) that is

Professor Emeritus of Surgery – NE Ohio Medical University, Rootstown, OH (Dr. Kavic).

Professor Emeritus of Surgery – University of Washington, Seattle, WA (Dr. Satava).

Address correspondence to: Michael S. Kavic, MD, FACS, Professor Emeritus of Surgery – Northeast Ohio Medical University, 431 Eastons Lane, Oakdale, PA 15071, USA. Telephone: 305-665-9959, E-mail: mkavic@SLS.org.

DOI: 10.4293/JSLS.2021.00010

based on Web of Science data. This citation metric was first proposed by Eugene Garfield in 1955 and subsequently developed to compare the quality of one journal to another journal.[4] As originally conceived, the JIF was intended as a device to assist librarians identify appropriate journals to purchase. It was not designed to gage the quality of research in an individual article.

The JIF is a journal level metric, not to be confused with article level metrics. The JIF is calculated by dividing the total number of citations that a journal receives in a two-year period by the number of articles it published in that same time frame. The JIF is commonly used to afford proof of the impact and legitimacy of a scientific journal in the surgical community.

But this metric, the JIF, has morphed into something quite different from that originally intended. The JIF is now regularly regarded as a surrogate for the impact of all the articles published in a particular journal, and, by extension, the impact of the authors of these articles. Some even take into consideration the JIF when evaluating an author's department and university.[1] Many now evaluate individual performance by the journal impact factor and the JIF may influence decisions made by university tenure and hiring committees as well as funding agencies.[5]

Certainly, the JIF was not designed for, and is not well suited for assessing quality of the massive amount of scientific literature available during the present era. It just does not have the scope or depth of penetrative analysis to provide all the information necessary to evaluate 21st century scientific literature. In addition, the JIF has a built-in time lag of two years and is patently not fast enough to keep up in an internet driven world.

To help fill the limitations in analytical power of the JIF and to further leverage citation data, other metrics have been developed. These include, but are not limited to: Eigenfactor, article influence score, CiteScore, source normalized impact per paper (SNIP), scmigao journal rank (SJR), and the h-index.[6]

Eigenfactor

Eigenfactor measures the influence of a journal based on whether it is cited within other reputable journals over five years. Eigenfactor calculations borrow their methodology from network theory and use an algorithm to rank the influence of journals according to the citations they receive. A citation from a highly cited journal is worth more than one from a journal less cited. The citations are also weighted by the length of the reference list they are from. Scores do not take journal size into account; therefore, larger journals tend to have larger Eigenfactors as they receive more overall citations.

An Eigenfactor calculation follows: number of citations in one year to content published in Journal Z in the previous five years (weighted), divided by the total number of articles published in Journal Z within the previous five years.

Article Influence Score

Article influence score is a measure of the average influence of a journal's articles in the first five years after publication. It looks like this: (0.01X Eigenfactor) divided by number of articles published in that journal over five years, divided by the number of articles published in all journals over five years.

CiteScore

CiteScore is the ratio of citations to research published and indexed in Scopus. It considers all content published in a journal, not just articles and reviews. The CiteScore calculation: Number of all citations recorded in Scopus in one year to content published in Journal Z in the last three years, divided by the total number of items published in Journal Z in the past three years.

Snip

Source normalized impact per paper (SNIP) is a journal level metric. It measures citations received against citations expected for the subject using Scopus data over a three-year period. It is calculated in this fashion: Journal citation count per paper, divided by citation potential in the field.

Scimago

Scimago journal rank (SJR) seeks to capture the effect of subject field, quality, and reputation of a journal on citations. It attempts to calculate the prestige of a journal by considering the value of the sources that cite it using Scopus data. The calculation follows: Average number of weighted citations in a given year to Journal Z, divided by the number of articles published Journal Z the previous three years.

h-Index

h-index is an author-level research metric. The h-index attempts to measure the production of a researcher and the citation impact of the author's publications. The calculation is as follows: Number of articles published which have received the same number of citations.

Despite the above, to evaluate and filter the deluge of information of the digital age has increasingly become a matter of major concern. At the present moment readers, clinicians and research investigators now evaluate the quality of scientific literature using three different methods: citation metrics (described above), usage metrics, and alterative metrics (altmetrics).[2] Bollen et al. suggest that the concept of scientific impact involves a multidimensional construct that cannot be adequately measured by any one single indicator.[1] They suggest that multiple measures are needed and currently available. These include: 1) citation metrics (described above), 2) usage metrics, and 3) alternative metrics (altmetrics).[2]

Usage Metric

Usage metric – an article level metric - is a comparatively new way to assess the impact of a scientific article. Usage data relates to the number of PDF downloads or HTML views an article or journal receives and provides an immediate reflection of a journal's reach.[6] A compelling reason to consider usage metrics is that data accrues immediately after publication, is readily collected, and available for analysis. This metric should be used in parallel with other citation-based metrics to more thoroughly measure a journal's performance.

Sheppard and Bollen have expanded the potential usefulness of usage data. They suggest developing a tool similar to the journal impact factor. The tool, a usage impact factor, would take advantage of the immediacy and breadth of usage data and eliminate the constraint of a two-year waiting period required for a JIF. The usage impact factor would be derived from averaging usage rates for articles published in a journal over a defined period of time similar to the citation-based JIF.[7,8]

There are, however, disadvantages to usage metrics. A major issue is that full-text downloads do not signify use of that information. There is no way to determine if downloads are put into clinical practice or that they are even read. Realistically then, full-text downloads should comprise one part of a mosaic of tools necessary to encompass the scope of material involved in transferring scientific knowledge. That downloads are part of the mosaic of transferring scientific knowledge is not in question, but the nature of the effect measured by downloads is not yet well defined.[9]

Altmetrics

Alternative metrics provide additional information on article usage by looking at the social activity around the journal article. They use qualitative and quantitative data (posts, blogs, Facebook, tweets, and other social media) alongside traditional citation-based metrics and usage-based metrics to provide insight into the attention, influence, and impact of academic research. The Altmetric Attention Score is the most common method of reporting on altmetrics.[6] The major concern with these social media platforms is verifying their authenticity. Recall the quotation which circulated at the origin of the internet: "On the internet, nobody knows you're a dog."[10]

All of the above metrics are represented by numbers. Numbers are an arithmetical value expressed as a figure, word, or symbol. Data, for our purposes, refers to information expressed as numbers. Because the same numbers, without context, can represent entirely different concepts, Carroll D. Wright in 1889, said that "figures do not lie," but qualified this with "liars will figure."[11]

Because of the inherent elasticity of numeric values and data interpretation, metrics that depend solely on numbers can be manipulated – the system can be gamed. In scientific publication, the JIF is the most widely used bibliometric index and it can be manipulated. Several stratagems have been identified that are used to inflate a JIF. These include:

1. Publishing large number of papers that are sure to get cited.
2. Increasing self-citation to the journal.
3. Publishing reviews which typically get more citations.
4. Publishing papers of questionable scientific value because of expectations they will be widely cited.[12]

To help counter deliberate manipulation of scientific information there has developed a set of obligations (ethical principles) that define right and wrong in scientific publication. Scientific ethics demands honesty and integrity in all aspects of the research process. These core principles include:

1. Honesty in reporting of scientific data;
2. Careful transcription and analysis of scientific results to avoid error;

3. Independent analysis and interpretation of results that is based on data and not on the influence of external sources;
4. Open sharing of methods, data, and interpretations through publication and presentation;
5. Sufficient validation of results through replication and collaboration with peers;
6. Proper crediting of sources of information, data, and ideas;
7. Moral obligation to society in general, and, in some disciplines, responsibility in weighing the rights of human and animal subjects.[13]

In the final analysis, it remains up to each clinician and academician to educate themselves on how to evaluate the scientific literature. Then, with the tools available, determine which articles to read and study. There are three different, commonly available metrics to evaluate the scientific literature – citation metrics (journal level metric), usage metrics (article level metric), and altmetrics (described above).

Journal level metrics, such as the JIF, do not speak to the quality of an individual scientific article in that journal.[14] There is currently no metric that specifically does so. But, it is reasonable to consider the number of times a full-text download of that article has occurred as indicative of the importance of that article to the scientific community. Full-text downloads of an article speak, at a certain level, to the intent of the researcher to read that article and learn from it. Usage metrics capture this aspect of the evaluation process. Sheppard and Bolen are correct regarding the value of usage data. It may be possible to combine the journal impact factor with a usage metric (full-text downloads) to come up with a meaningful measure that would be of value to the scientific, research, and clinical communities. Authors, institutions, governing agencies, and the scientific publishing community should insist this usage tool (full-text downloads) be more fully developed and incorporated in the evaluation of scientific literature.

References:

1. Bollen J, Van de Sompel H, Hagberg A, Chute R. A principal component analysis of 39 scientific impact measures. *PLos One*. 2009;4(6):e6022. Accessed January 19, 2021.

2. DiBartola SP, Hinchcliff KW. Metrics and the scientific literature: deciding what to read. *J Vet Intern Med*. 2017;31(3):629–632.

3. Levy S. Hackers: Heroes of the computer revolution. 1984. Anchor Press/Doubleday. ISBN: 0-385-19195-2.

4. Garfield E. Journal impact factor: a brief review. *Canadian Medical Journal*. 1999;161:979–980.

5. Callaway E. Beat it, impact factor! Publishing elite turns against controversial metric. *Nature*. 2016;535(7611):210–1.

6. Cuschieri S. WASP (Write a Scientific Paper): Understanding research metrics. *Early Hum Dev*. 2018;118:67–71.

7. Shepherd PT. The feasibility of developing and implementing journal usage factors: a research project sponsored by UKSG. *Serials*. 2007;20(2):117–123.

8. Bollen J, Van de Sompel H. Usage impact factor: the effects of sample characteristics on usage-based impact metrics. *J Am Soc Inf Sci*. 2008;59(1):136–149.

9. Kurtz MJ, Heineken EA. Measuring metrics – a 40-year longitudinal cross-validation of citations, downloads, and peer review in astrophysics. *Journal of the Association for Information Science and Technology*. 2017;68:695–708.

10. Jones CR. "Nobody knows you're a dog". In Land Ray, Bayne Sian (eds.). *Education in Cyberspace. New York: Routledge*. 2004.

11. Wright Carroll D. 1890 (1889 June 25, date of opening remarks at the convention), Joint Documents of the State of Michigan for the Year 1889: Volume III Part I., Seventh Annual Report of the Bureaus of Labor and Industrial Statistics Issued February 1, 1890, Page 311, Thorp Darius D.: State Printer and Binder, Lansing Michigan.).

12. Ioannidis JP, Thombs BD. A user's guide to inflated and manipulated impact factors. *Eur J Clin Invest*. January 2021;49(9) Epub 2019 Jul 1.

13. Carpi A, Egger AE. The Culture of Science. "Scientific Ethics". *Visionlearning*. 2009;P05-2(5). https://www.visionlearning.com/en/librar/Process-of-Science/49/Scientific-Ethics/161. Accessed January 12, 2021.

14. Raff J. The San Francisco Declaration of Research Assessment. *Biology Open*. 2013;2(6):533–534. https://sfdora.org/read. Accessed January 16, 2021.