Research assessment: Review of methodologies and approaches

Henk F. Moed Dr

Gali Halevi Dr
Elsevier

Follow this and additional works at: https://www.researchtrends.com/researchtrends

Research Assessment

Review of methodologies and approaches

Dr. Henk Moed and Dr. Gali Halevi

The assessment of scientific merit and individuals has a long and respectable history which has been demonstrated in numerous methods and models utilizing different data sources and approaches (1, 2). The proliferation and increasing availability of primary data has created the ability to evaluate research on many levels and degrees of complexity, but has also introduced some fundamental challenges to all who are involved in this process, including evaluators, administrators and researchers, and others (3).

Evaluative methods are used on several levels within the scientific world: (1) Institutional (including departmental) level, (2) Program level, and (3) Individual level. Each of these levels has its own objectives and goals; for example:

Institutional evaluation is being used in order to establish accreditation, define missions, establish new programs and monitor the quality of an institute's research activities among others. The types of evaluative results can be seen in the ranking systems of universities, which at present are produced at both regional and international levels, based on different criteria (4). Institutional evaluations are performed based on prestige measures derived from publications, citations, patents, collaborations and levels of expertise of the individuals within the institution.

Program level evaluations are performed in order to measure the cost-benefit aspects of specific scientific programs. These are usually based on discovering the linkage between the investment made and the potential results of the program (5). Within this realm we find measures developed for technology transfer capabilities and commercialization potentialities of the program, among others (6).

Finally an **individual evaluation** is mainly performed for purposes of promotion and retention of individuals and is done at specific times in a researcher's career. Individual assessment methods rely mainly on counts of publications or citations (7). In the past few years, with the advent of social media, we have seen an increase in the use of measures based on mentions in social media sites such as blogs, Facebook,

LinkedIn, Wikipedia, Twitter, and others, which are labelled as sources of "altmetrics" and include news outlets as well (8). The data used for each of these evaluation goals, whether they measure a publication's impact in social or economic terms or both, varies by the method chosen in each case.

Evaluative Indicators Based on the different methodologies and approaches, several indicators, aimed at quantifying and benchmarking the results of these evaluative methods, have emerged through the years. Table 1 summarizes some of the main indicators used today in research evaluation.

Big data and its effect on evaluation methods

Big data refers to a collection of data sets that is so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The advent of supercomputers and cloud computing which are able to process, analyze and visualize these datasets also has an effect on evaluation methods and models. While a decade ago scientific evaluation relied mainly on citations and publication counts, most of which were even done manually, today these data are not only available digitally but can also be triangulated with other data types (9). Table 2 depicts some examples of big datasets that can be combined in a bibliometric study to investigate different phenomena related to publications and scholarly output. Thus, for example, publication and citation counts can be triangulated with collaborative indicators, text analysis and econometric measures to produce a multi-level view of an institution, program or an individual. Yet, the availability and processing capabilities of these large datasets does not necessarily mean that evaluation becomes simple or easy to communicate. The fact of the matter is that as they become more complex, both administrators and evaluators find it difficult to reach consensus as to which model best depicts the productivity and impact of scientific activities. These technological abilities are becoming a breeding ground for more indices, models and measures and while each may be valid and grounded in research, they present a challenge in deciding which are best to use and in what setting.

Type of Indicator	Description	Main Uses	Main Challenges
Publications – Citations	Methods involving counts of the number of publications produced by the evaluated entity (e.g. researcher, department, institution) and the citations they receive.	Measuring the impact and intellectual influence of scientific and scholarly activities including: • Publication impact • Author impact • Institution/department impact • Country impact	Name variations of institutions and individuals make it difficult to count these correctly. Limited coverage or lack of coverage of the database selected for the analysis can cause fundamental errors. Some documents such as technical reports or professional papers ("grey literature") are usually excluded from the analysis due to lack of indexing and thus, in certain disciplines, decrease the accuracy of the assessment. Differences between disciplines and institution reading behaviors are difficult to account for.
Usage	Methods that aim to quantify the number of times a scholarly work has been downloaded or viewed.	Indicating works that are read, viewed or shared as a measure of impact. Enabling authors to be recognized for publications that might be less cited but heavily used.	Incomplete usage data availability across providers leads to partial analysis. Differences between disciplines and institution reading behaviors are difficult to account for. Content crawling and automated downloads software tools that allow individuals to automatically download large amounts of content, which doesn't necessarily mean that it was read or viewed. Difficult to ascertain whether downloaded publications were actually read or used.
Social (Altmetrics)	Methods that aim to capture the number of times a publication is mentioned in blogs, tweets or other social media platforms such as shared reference management tools.	Measuring the mentions of a publication in social media sites, which can be considered as citations and usage and thus indicate the impact of research, an individual or institution.	A relatively new area with few providers of social media tracking. The weight given to each social media source is different from one provider to the other thus leading to different "impact" scores.
Patents	Measures the number of patents assigned to an institution or an individual. Identification of citations to basic research papers in patents as well as patents that are highly cited by recently issued patents.	Attempting to provide a direct link between basic science and patents as an indication of economic, social and/or methodological contribution.	Incomplete and un-standardized references and names limit the ability of properly assigning citations and patents to individuals or institutions. Patenting in countries other than where the institution or individual originates from is problematic for impact analysis. Lack of exhaustive reference lists within the patents limits the analysis.
Economic	Measures the strengths between science and its effect on industry, innovation and the economy as a whole.	Providing technology transfer indicators. Indicating patentability potentialities of a Research project. Providing cost-benefit measures.	The statistical models used are complex and require deep understanding of the investment made but also of the program itself. Long term programs are more difficult to measure as far as the cost-benefit is concerned. Requires expertise not only in mathematics and statistics but also in the field of investigation itself.
Networks	Calculates collaborations between institutions and individuals on a domestic and global scale. Institutions and individuals that develop and maintain a prolific research network are not only more productive but also active, visible and established.	Enabling the tracking of highly connected and globally active individuals and institutions. Allowing benchmarking to be performed by evaluators by comparing collaborating individuals and institutions to each other.	Affiliation names as mentioned in the published papers are not always standardized, thus making them difficult to trace. Education in a different country which might not have resulted in a publication cannot be measured, thus making this particular aspect of expertise building impossible to trace.

Table 1: Types of evaluative indicators.

Conclusions

As shown by this brief review, research assessment manifests itself in different methodologies and indicators. Each methodology has its strengths and limitations, and is associated with a certain risk of arriving at invalid outcomes. Indicators and other science metrics are essential tools on two levels: in the assessment process itself, and on the Meta level aimed to shape that process. Yet, their function on these two levels is different. On the first they are tools in the assessment of a particular unit, e.g. a particular individual researcher, or department, and may provide one of the foundations of evaluative statements about such a unit. On the second level they provide insight into the functionality of a research system as a whole, and help draw general conclusions about its state, assisting in drafting policy conclusions regarding the overall objective and general set-up of an assessment process.

Closely defining the unit of assessment and the evaluative methodologies to be used can provide a clue as to how peer review and quantitative approaches might be combined. For instance, the complexity of finding appropriate peers to assess all research groups in a broad science discipline in a national research assessment exercise may urge the organizers of that exercise to carry out a bibliometric study first and decide on the basis of its outcomes in which specialized fields or for which groups a thorough peer assessment seems necessary.

As Ben Martin pointed out in his 1996 article (10), this is true not only for metrics but also for peer review. It is the task of members from the scholarly community and the domain of research policy to decide what are acceptable "error rates" in the methodology and indicators being used, and whether its benefits prevail over its detriments. Bibliometricians and other science and technology analysts should provide insight into the uses and limits of various types of metrics, in order to help scholars and policy makers to carry out such a delicate task.

Combined datasets	Studied phenomena	Typical research questions
Citation indexes and usage log files of full text publication archives	Downloads versus citations; distinct phases in the process of processing scientific information	What do downloads of full text articles measure? To what extent do downloads and citations correlate?
Citation indexes and patent databases	Linkages between science and technology (the science–technology interface)	What is the technological impact of a scientific research finding or field?
Citation indexes and scholarly book indexes	The role of books in scholarly communication; research productivity taking scholarly book output into account	How important are books in the various scientific disciplines, how do journals and books interrelate, and what are the most important book publishers?
Citation indexes (or publication databases) and OECD national statistics	Research input or capacity; evolution of the number of active researchers in a country and the phase of their career	How many researchers enter and/or move out of a national research system in a particular year?
Citation indexes and full text article databases	The context of citations; sentiment analysis of the scientific-scholarly literature	In what ways can one objectively characterize citation contexts? And identify implicit citations to documents or concepts?

Table 2: Compound Big Datasets and their objects of study. Source: Research Trends, Issue 30, September 2012 (9).

References:

1. Vale, R. D. (2012). "Evaluating how we evaluate", Molecular Biology of the Cell, Vol. 23, No. 17, pp. 3285-3289.

2. Zare, R. N. (2012). "Editorial: Assessing academic researchers", Angewandte Chemie - International Edition, Vol. 51, No. 30, pp. 7338-7339.

3. Simons, K. (2008). "The misused impact factor", Science, Vol. 322, No. 5899, pp. 165.

4. O'Connell, C. (2013). "Research discourses surrounding global university rankings: Exploring the relationship with policy and practice recommendations", Higher Education, Vol. 65, No. 6, pp. 709-723.

5. Guido M. Imbens and Jeffrey M. Wooldridge (2008). "Recent Developments in the Econometrics of Program Evaluation", The National Bureau of Economic Research Working Papers Series. Available at: http://www.nber.org/papers/w14251

6. Arthur, M. W., & Blitz, C. (2000). "Bridging the gap between science and practice in drug abuse prevention through needs assessment and strategic community planning", Journal of Community Psychology, Vol. 28, No. 3, pp. 241-255.

7. Lee, L. S., Pusek, S. N., McCormack, W. T., Helitzer, D. L., Martina, C. A., Dozier, A. M., & Rubio, D. M. (2012). "Clinical and Translational Scientist Career Success: Metrics for Evaluation", Clinical and translational science, Vol. 5, No. 5, pp. 400-407.

8. Taylor, M. (2013). "Exploring the boundaries: How altmetrics can expand our vision of scholarly communication and social impact", Information Standards Quality, Vol. 25, No. 2, pp. 27-32. Available at: http://www.niso.org/publications/isq/2013/v25no2/taylor/

9. Moed, H.F., (2012). "The Use of Big Datasets in Bibliometric Research", Research Trends, Issue 30, September 2012. Available at: http://www.researchtrends.com/issue-30-september-2012/the-use-of-big-datasets-in-bibliometric-research/

10. Martin, B.R., (1996). "The use of multiple indicators in the assessment of basic research", Scientometrics, Vol. 36, No. 3, pp. 343-362.