The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects

Stefanie Haustein and Vincent Larivière

Abstract Researchers are used to being evaluated: publications, hiring, tenure and funding decisions are all based on the evaluation of research. Traditionally, this evaluation relied on judgement of peers but, in the light of limited resources and increased bureaucratization of science, peer review is getting more and more replaced or complemented with bibliometric methods. Central to the introduction of bibliometrics in research evaluation was the creation of the Science Citation Index (SCI) in the 1960s, a citation database initially developed for the retrieval of scientific information. Embedded in this database was the Impact Factor, first used as a tool for the selection of journals to cover in the SCI, which then became a synonym for journal quality and academic prestige. Over the last 10 years, this indicator became powerful enough to influence researchers' publication patterns in so far as it became one of the most important criteria to select a publication venue. Regardless of its many flaws as a journal metric and its inadequacy as a predictor of citations on the paper level, it became the go-to indicator of research quality and was used and misused by authors, editors, publishers and research policy makers alike. The h-index, introduced as an indicator of both output and impact combined in one simple number, has experienced a similar fate, mainly due to simplicity and availability. Despite their massive use, these measures are too simple to capture the complexity and multiple dimensions of research output and impact. This chapter provides an overview of bibliometric methods, from the development of citation indexing as a tool for information retrieval to its application in research evaluation, and discusses their misuse and effects on researchers' scholarly communication behavior.

1. Introduction

The evaluation of researchers' work and careers, which traditionally relied on peer review, is increasingly substituted or influenced by publication output and citation impact metrics (Seglen, 1997b; Rogers, 2002; Cameron, 2005). Bibliometric indicators are more and more applied by governments and other funding organization mainly because of their large-scale applicability, lower costs and time as well as their perceived objectivity. The goal is to optimize research allocations and make funding both more efficient and effective (Moed, 2005; Weingart, 2005). Bibliometrics and citation analysis go back to the beginning of the twentieth century, when they were used by librarians as a tool for journal collection development. Although mentioned much earlier by Otlet (1934) in French (*bibliométrie*), it is Pritchard (1969) who is mostly associated with coining the term *bibliometrics* as a method

S. Haustein
École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal
Montréal, Canada
e-mail: stefanie.haustein@umontreal.ca

V. Larivière
École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal
Montréal, Canada
e-mail: vincent.lariviere@umontreal.ca

"to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and analyzing the various facets of written communication." (Pritchard 1969, pp. 348-349)

However, it is only with the creation of the Science Citation Index in the early 1960s that bibliometrics became a method that could be massively applied to analyze patterns of scholarly communication and evaluate research output. Over the last 20 years, the increasing importance of bibliometrics for research evaluation and planning led to an oversimplification of what scientific output and impact were which, in turn, lead to adverse effects such as *salami publishing*, *honorary authorships*, *citation cartels* and other unethical behavior to increase one's publication and citation scores, without actually increasing one's contribution to the advancement of science (Moed, 2005).

The goal of this chapter is to inform the reader about bibliometrics in research assessment and explain possiblilities and limitations. The chapter starts with a brief historic summary of the field and the developments of its methods, which provides the context in which the measurement of scholarly communication developed. An overview of indicators and their limitations is then provided, followed by their adverse effects and influence on researchers' scholarly communication behavior. The chapter concludes by summarizing the possibilities and limitations of bibliometric methods in research evaluation.

2. Development of Bibliometrics and Citation Analysis

Bibliometric analyses are based on two major units: the scientific publication as an indicator of research output and citations received by them as a proxy of their scientific impact or influence on the scholarly community. Early bibliometric studies, at that time referred to as *statistical bibliography*, were mostly applied to investigate scientific progress and later library collection management. For example, Cole and Eales (1917), which can be considered the first bibliometric study, examined the scientific output of European countries in anatomy research based on literature published between 1543 and 1860. By defining publications as the main unit of measurement to assess scientific activity in certain research areas, they laid out the basis for future bibliometric studies (De Bellis, 2009). Ten years later, Gross and Gross (1927) were the first ones to carry out a citation analysis of journals, in search for an objective methods for the management of their library collection. Pressured by limited budgets and physical space in libraries opposed to the ever-growing volume of scholarly documents published, they extracted citations to journals from 3,663 references listed in the 1926 volume of the Journal of the American Chemical Society, thus compiling a list of most cited journals to subscribe to. In doing so they equated citations received by journals with their importance in the disciple, setting the stage for citation analysis in both library collection management and research evaluation. Bradford (1934) further influenced librarians and collection management through his famous law of scattering, stating that the majority of documents on a given subject are published in a small number of core journals. Together with Lotka's (1926) law on the skewed distributions of papers per author, Zipf's (1949) law on word frequencies in texts, as well as Price's (1963) work on the exponential growth of science, Bradford formed the basis of the mathematical foundations of the field of bibliometrics. It did, however, take the development of a global and interdisciplinary citation index, i.e. Garfield's Science Citation Index (SCI), for bibliometric methods – and citation analysis as its key aspect – to enter into the area of research policy and evaluation. Citation indexes were the key to evaluative bibliometrics and research evaluation because they provided the database and made global and large-scale analysis feasible. The development of the Institute of Scientific Information (ISI) and the SCI gave rise to both the practical application of bibliometrics in research evaluation and information retrieval and theoretical and empirical research of citation analysis and bibliometric indicators.

2.1. The Science Citation Index

After World War II it was believed that economic growth and scientific progress were intertwined and the latter could be controlled and steered towards specific goals, resulting in the era of big science and hyperspecialization (De Bellis, 2009). It is in this context that citation indexing developed as a means to cope with the flood of scientific literature. With the growth of publication output, the scientific landscape had become complex and the amount of literature unmanageable. Garfield's citation indexes aimed at making the information overload manageable creating a "World Brain" (Garfield, 1964) of scientific information through automatic indexing based on references. Garfield adopted this method from Shepard's Citation, a citatory service in the field of law established in 1873 to keep track of the application of legal decisions (Garfield, 1979). Citations were assumed to be the better descriptors and indexing terms as symbols of a document's content than natural language such as terms derived from document titles (Garfield, 1964). Garfield believed that the community of citing authors outperformed indexers in highlighting cognitive links between papers especially on the level of particular ideas and concepts (Garfield, 1983, p. 9), an approach resembling the pheonomenon known today as *crowdsourcing*. This is the theoretical foundation that citation indexing is based on.

As a multidisciplinary citation index, the SCI was initially developed for information retrieval and not for research evaluation. Along these lines, the Impact Factor was initially developed by Garfield as a tool to select the most relevant journals for coverage in ISI's SCI, with a particular focus on cost efficiency. As the mean number of citations received in one year by papers published in a journal during the two previous years, the Impact Factor selects the most cited journals regardless of output size (Garfield, 1972).

Garfield's law of concentration, a further development of Bradfield's law of scattering combining all fields of science, showed that the majority of cited references referred to as few as 500 to 1,000 journals, justifying a cost-efficient coverage approach (Garfield, 1979). Based on the Impact Factor, 2,200 journals had been identified by 1969 as "the world's most important scientific and technical journals" (Garfield, 1972, p. 471) and became fully indexed by the SCI. The ISI citation indexes fostered further developments of the field of bibliometrics in general and of citation analysis in particular, both empirically and theoretically. By enabling large-scale publication and citation analysis of different entities from micro (author) to macro (country) level, the SCI provided the basis for quantitative research evaluation. Garfield himself underlined the potential of citation analysis in research evaluation and outlined the usefulness of the Impact Factor for librarians, editors and individual scientists (Garfield, 1972; Moed, 2005).

2.2. Theory of Publication and Citation Analysis

Original research results are typically formally communicated through publications. Thus, publications can be regarded as proxies of scientific progress at the research front (Moed, 2005). They do, however, not capture the entire spectrum of scientific activity. In most of the medical and natural sciences, the journal article is the main publication format used by researchers to disseminate and communicate their findings to the research community, claim priority of findings and make them permanent. Peer-review and editorial work ensure a certain level of quality control, and details on the methods provide the means for colleagues to replicate the findings. Given this central importance of articles for scholarly communication, sociological research has considered that, by counting papers, we obtain an indicator of research activity.

Citation analysis is based on the assumption that a document referenced in a subsequent paper marks the intellectual influence of the cited document on the citing paper. Since the number of cited items is usually restricted by the length of the publication, a reference lists should not be considered a complete list but a selection of the most influential sources related to a piece of work (Small, 1987). The number of citations received is thus assumed to reflect the influence or scientific impact of scholarly documents and mark their contribution to progress and advancement of science. This assumption is based on the argument by sociologist Robert K. Merton (1977, pp. 54-55) that "if one's work is not being noticed and used by others in the system of science, doubts of its value are apt to rise." Merton's normative approach regards

citations as the "pellets of peer recognition" (Merton, 1988, p. 620) within the scientific reward system, a symbol of acknowledging the knowledge claim of the cited source. The work by Merton's students Stephen and Jonathan Cole (Cole & Cole, 1973) and Harriet Zuckerman (Zuckerman, 1987), which analyzed Merton's theories from an empirical perspective, have shown positive but not perfect correlations between citation rates and qualitative judgment by peers, thus providing an early framework for the use of bibliometrics in assessing, albeit imperfect, scientific influence (Moed, 2005).

2.3. Bibliometrics and Peer Review

Peer review is the most important instrument when it comes to judging or ensuring the quality of scientists or of their work. It is applied at each level of a researchers' career, from submitted manuscripts to the evaluation of grant proposals, as well as to their suitability for academic positions or scholarly awards. Based on Merton's set of norms and values conveyed by the *ethos of science*, in a perfect setting peer review should be based entirely on the scientific quality and disregard any personal interests (Merton, 1973). In reality, judgements could be influenced by prejudices and conflicts of interest of the referee, are sometimes inconsistent or often contradict each other (Martin & Irvine, 1983). In addition, peer review is time and cost intensive. This has generated, for some, the need for faster and more cost-efficient methods. Studies (e.g., Cole & Cole, 1967; Martin & Irvine, 1983; Rinia et al., 1998; Norris & Oppenheim, 2003) correlating peer judgement with citations found positive but not perfect correlations, indicating that the two approaches reflect similar but not identical assessments. Given the limitations of both methods, none of them leads to a perfect, unbiased quality judgement. In the evaluation of research output, peer review and bibliometrics do thus not replace each other but are best used in combination.

3. Bibliometric Analyses

3.1. Basic Units and Levels of Aggregation

Aggregation levels of bibliometric studies range from micro (author) to macro (countries) with different kinds of meso level inbetween such as institutions, journals or research fields and subfields. Regardless of the level of aggregation, publication activity of particular entities is determined through the author and author addresses listed in the article (De Lange & Glänzel, 1997). A paper published by author A at Harvard University and author B at the University of Oxford in the Astrophysical Journal would thus count as a publication in astrophysics for authors A and B on the micro, Harvard and Oxford on the meso and the US and the UK on the macro level. Since co-publications serve as a proxy of scientific collaboration, the same publication would provide formal evidence of authors A and B, Harvard and Oxford and the US and UK collaborating in astrophysics research. Publications can be counted fully, i.e., each participating unit is credited with one publication, or fractionally, assigning an equal fraction of the paper to each entity (Price, 1981), that is 0.5 to each of the two entities per aggregation level in the example above. The latter is particularly helpful to compare scientific productivity of research fields with different authorship patterns. While single-authored papers are quite common in the humanities, the list of authors in experimental physics can include more than a hundred names, because research is carried out at large international facilities (Larivière, Gingras and Archambault, 2006). The number of authors per article and international collaboration is increasing over time for all fields of research (Abt, 1992; De Lange & Glänzel, 1997). Hence, the probability that a researcher contributes to a paper is, thus, very different from one field to another.

Similarly, different fields have different rules when it comes to authorship (Pontille, 2004; Birnholtz, 2006). For example, the act of writing the paper is central to authorship in the social sciences and humanities, while, in the natural and medical sciences, the data analysis plays an important role. As a consequence, a research assistant who analyzed the data in sociology would not typically be an author of the paper – and thus be included in the bibliometrics measures – while in natural and medical sciences, this task could lead to authorship. This is exemplified by the very large number of authors found in high-

energy physics, where all members of the research team – which can amount to several hundreds – will sign the paper, even several months after they left the experiment (Biagioli, 2003). Thus, the adequation between 'research activity' and bibliometrics depends on the discipline.

3.2. Data Quality

Another aspect that is essential in bibliometric studies is the quality of data. This involves the selection of a suitable database and cleaning of bibliographic metadata. Author and institution names come in many different forms including first names and intitials, abbreviations and department names or spelling errors and they may change over time (*synonymy problem*). On the other hand, the same name might refer to more than one person or department (*homonymy problem*). Disambiguation and cleaning author names and institutions is fundamental to computing meaningful bibliometric indicators used in research evaluation[1]. It is most successful if either publication lists are verified by authors or on a larger scale if cleaning is carried out by experts supported by powerful rule-based algorithms computing the probabilities based on similarities of document metadata including author addresses, co-authors, research fields and cited references (Moed, 2005; Smalheiser & Torvik, 2009). It should be assumed that the individual author would be able to identify his or her own publications. As the authors themselves should know best which papers they have written, a registry with unique and persistent author identifiers which researchers can link with their publications, would solve author homonymy and synonymy problems. ResearcherID[2] represents such a registry within Web of Science and ORCID[3] has recently been launched as a platform-independent and non-profit approach to identify and manage records of research activities (including publications, datasets, patents etc.). For such a system to fully replace author name disambiguation for evaluative bibliometrics, it would have to be based on the participation of the entire population of authors during the period under analysis including those that are no longer active in the field or even alive. As this is not very likely such registries can be used to support the disambiguation process but cannot entirely replace data cleaning in bibliometric analyses.

3.3. Bibliometric Indicators

Provided a cleaned dataset, entitites such as authors, institutions or countries can be compared regarding their publication activity and citation impact using bibliometric indicators. Among those frequently used in research evaluation, one can distinguish between *basic* and *normalized* metrics, *time-based* and *weighted indicators* are other kinds of citation indicators. Since they are more common in journal evaluation – e.g., cited half-life (Burton & Kebler, 1960) or Eigenfactor metrics (Bergstrom, 2007) and SCImago journal rank (Gonzalez-Pereira, 2010) – than in research assessment of authors, institutions and countries, they are not included in the following overview for reasons of space.

3.3.1. Basic Indicators

Basic or simple bibliometric indicators include the number of publications and citations, which are size-dependent measures. Mean citation rates try to account for output size by dividing the total number of citations received by an entity by the number of its publications. Basic citation rates can be limited to certain document types, include or exclude or self-citations or have different citation and publication windows. That is, they can be calculated synchronously − citations received in one year for documents published during previous years − or diachronously − documents published in one year cited in subsequent years (Todorov & Glänzel, 1988). Without fixed publication and citation windows or accurate normalization for publication age, an older document has higher probabilities of being cited.

---

[1] It should be mentioned that, despite the fact that bibliometrics should not be used alone for research evaluation, individual level disambiguation is ofted needed in order to assess groups of researchers.
[2] http://www.researcherid.com
[3] http://orcid.org

Basic publication and citation indicators are influenced by different publication patterns of disciplines and also by the size or age of the measured entity. Using basic instead of normalized (see 3.3.2) metrics, a researcher from the medical sciences would thus seem more productive and to have higher citation impact than a mathematician, because medical scientists contribute to more papers and their papers contain a larger number of and more recent references than those in mathematics. Comparing the publication output and citation impact of authors, institutions, journals and countries without an accurate normalization is thus like comparing apples with oranges. A university with a large medical department would always seem more productive and impactful than those without.

Mean citation rates are the most commonly used size-independent indicator of scientific impact. Due to the highly skewed distribution of citations per paper – as a rule of thumb, 80% of citations are received by 20% of documents and many are never cited, especially in the humanities (Larivière, Gingras and Archambault, 2008) – the arithmetic mean is, however, not a very suitable indicator since other than in a normal distribution it is not representative of the majority of documents (Seglen, 1992). The median has been suggested as a more appropriate due to its robustness (Calver & Bradley, 2009), but since it disregards the most frequently cited document, it cannot fully represent the citation impact of a set of papers. Providing the standard deviation with the mean and additional distribution-based indicators such as citation percentiles, for example the share of top 1% or 5% highly cited papers as a measure of excellence, seem more appropriate (Tijssen, Visser, & van Leeuwen, 2002; Bornmann & Mutz, 2011).

Journal Impact Factor As noted above, the Impact Factor was developed out of the need to select the most relevant journals to include in the SCI regardless of output size:

> "In view of the relation between size and citation frequency, it would seem desirable to discount the effect of size when using citation data to assess a journal's importance. We have attempted to do this by calculating a relative impact factor – that is, by dividing the number of times a journal has been cited by the number of articles it has published during some specific period of time. The journal impact factor will thus reflect an average citation rate per published article." (Garfield, 1972, p. 476)

The journal Impact Factor is a certain type of mean citation rate, namely a synchronous one based on citations received in year *y* by papers published in the two previous years, i.e., *y-1* and *y-2*. As such the above-mentioned limitations of arithmetic means to represent non-normal distributions apply (Todorov & Glänzel, 1988; Seglen, 1992). An extreme example of the susceptibility of the Impact Factor to single highly-cited papers is that of Acta Crystallographica A, which increased 24-fold from 2.051 in 2008 to 49.926 in 2009 because of a software review of a popular program to analyze crystalline structure cited 5,868 times in 2009 (Haustein, 2012).

In addition to being a mean citation rate, the Impact Factor has other limitations and shortcomings. It includes articles, reviews and notes as publication types while citations to all document types are considered, leading to an asymmetry between numerator and denominator (Moed & van Leeuwen, 1995; Archambault & Larivière, 2009). This asymmetry has led journal editors to "optimize" their journals' publication behavior (see section 4). Another shortcoming of the journal Impact Factor is its short citation windows, which goes back to convenience and cost-efficiency decisions made in the early days of the SCI (Martyn & Gilchrist, 1968; Garfield, 1972). Garfield (1972) found that the majority of citations are received within the first two years after publication. For some disciplines two years are not long enough to attract a significant number of citaitons, thus leading to large distortions (Moed, 2005). Since its 2007 edition, the Journal Citation Report (JCR) includes a five-year Impact Factor but the two-year version remains the standard. The asymmetry between numerator and denominator, which was caused by computational limitations in the 1960s and could easily be solved by document-based citation matching, however, still exists.

H-index Introduced by a researcher outside the bibliometric community physicist Jorge E. Hirsch, the h-

index has had enormous impact on the scholarly and bibliometric community (Waltman & van Eck, 2012) due to its attempt to reflect an author's publication output and citation impact with one simple integer. Although initially perceived as its strength, the oversimplification of the two orthogonal dimensions of publications and citations (Leydesdorff, 2009), is actually its greatest weekness. Hirsch (2005) defined the h-index of an author as follows:

> "A scientist has index *h* if *h* of his or her *Np* papers have at least *h* citations each and the other (*Np h*) papers have *h* citations each." (Hirsch, 2005, p. 16569)

In other words, for any set of papers ranked by the number of citations, *h* indicates the number of papers for which the number of citations is equal to or higher than the corresponding ranking position, i.e., an author has an h-index of 10, if 10 of his papers were each cited at least 10 times. The set of documents from the first to the hth position are part of the so-called h-core. The indicator does not take into account the total number of publications or citations, so that two researchers with the same h-index could differ completely in terms of productivity and citation impact as long as they both published *h* papers with *h* citations.

Besides the fact that the metric tries to oversimplify a researcher's impact and is size-dependent, the h-index is also inconsistent. For example, if two authors gain the same number of citations even for the same co-authored document, their h-indexes increase by 1 only if the additional citations move the paper up in the ranking from a position outside to inside the h-core. Thus an identical citation increase even for the same paper can lead to different outcomes for two reseachers (Waltman & van Eck, 2012). Similarly, given that the maximum of the h-index is the entity's number of publication, the h-index is more strongly determined by the number of publications rather than the number of citations. Thus, the h-index cannot be considered as a valid indicator of research productivity and impact.

3.3.2. Normalized Indicators

As mentioned above, certain biases occur caused by differences in publication behavior between research fields, publication growth and speed, different document types, time frames and/or database coverage. To allow for a fair comparison of universities or researchers active in different subject areas, normalized citation indicators try to counterbalance these biases. A full normalization of biases is a difficult and so far not yet entirely solved task due to the complexity of processes envolved in scholarly communication.

The most commonly used field-normalized indicators are based on the so-called *a posteriori, ex post facto* or *cited-side* normalization method, where normalization is applied after computing the actual citation score (Glänzel et al., 2011; Zitt, 2010). The actual or observed citation value of a paper is compared with the expected discipline-specific world average based on all papers published in the same field in the same year and in some cases the same document type. Each paper thus obtains an observed vs. expected ratio, which above 1 indicates a relative citation impact above world average and below 1 the opposite (Schubert & Braun, 1996). An average relative citation rate for a set of papers – for example, all papers by an author, university or country – is computed by calculating the mean of the relative citation rates, i.e., observed vs. expected citation ratios, of all papers (Gingras & Larivière, 2011; Larivière & Gingras, 2011). As observed citation impact is compared to field averages the cited-side normalization relies on a pre-established classification system to define a benchmark. A paper's relative impact thus depends on and varies with different definitions of research fields. Usually classification systems are journal based thus causing problems particularly for inter- and multidisciplinary journals. An alternative to the cited-side method is the *citing-side* or *a priori* normalization, which is independent of a pre-established classification system because the citation potential is defined through citing behavior, i.e. the number of references (Leydesdorff & Opthof, 2010a; Moed, 2010; Zitt, 2010; Waltman & van Eck, 2010b).

Although normalized indicators are the best way to compare citation impact of different entities in a fair way, the complex structures of scholarly communication are difficult to capture in one indicator of

citation impact. It is thus preferable to triangulate methods and use normalized mean citation rates in combination with distribution-based metrics to provide a more complete picture.

4. Misuse and Adverse Effects of Bibliometrics

Adverse effects, misapplication and misuse of bibliometric indicators can be observed on the individual as well as the collective level. Researchers and journal editors look for ways to optimize or manipulate the outcomes of indicators targeted at assessing their success, resulting in changes of publication and citation behavior, while universities and countries reward publishing in high-impact journals. The more bibliometric indicators are used to evaluate research outputs and as a basis for funding and hiring decisions, the more they foster unethical behavior. The higher the pressure, the more academics are tempted to take shortcuts to inflate their publication and citation records. Misapplication and misuse of indicators such as the cumulative Impact Factor are often based on the uninformed use of bibliometric methods and data sources and develop in an environment where any number beats no number. The most common adverse effects and misuses of bibliometrics are described in the following.

Publishing in Journals That Count The importance of the Web of Science and the journal Impact Factor have led researchers to submit their papers to journals which are covered by the database and preferably to those with the highest Impact Factors, sometimes regardless of the audiences (Rowlands & Nicholas, 2005). For example, results from the Research Assessment Exercise in the UK show that the share of publications in journals in the social sciences increased from 49.0% in 1996 to 75.5% in 2008.[4] At the same time, more and more publications are published in English instead of national languages (Engels, Ossenblok & Spruyt, 2012; Hicks, 2013). Although this can be seens as a positive outcome, it has adverse effects on the research system as it can lead to a change in scholars' research topics, especially in the social sciences and humanities. More specifically, given that journals with higher Impact Factors are typically Anglo-American journals that focus on Anglo-American research topics, scholars typically have to work on more international or anglo-american topics in order for their research to be published in such journals and, as a consequence, perform much less research on topics of local relevance. For instance, at the Canadian level, the percentage of papers authored by Canadian authors that have "Canada" in the abstract drops from 19% in Canadian journals to 6% in American journals (compilation by the authors based on the Web of Science).

Salami Publishing and Self-Plagiarism Increasing the number of publications by distributing findings across several documents is known as *salami slicing, duplicate publishing* or *self-plagiarism*. This practice, where one paper is sliced into many small pieces to increase the number of countable output resulting in the smallest publishable unit, or when results from previously published papers are republished without proper acknowledgement, is regarded as unethical as it distorts scientific progress and wastes the time and resources of the scientific community (Martin, 2013). The extent of such practices has been found to range between 1% and 28% of papers depending on the level of the plagiarism (Larivière and Gingras, 2010), which can span from the reuse of some data or figures to the duplication of the entire document. With the increase of the use of quantitative methods to assess researchers, we can expect such duplication to become more important and, thus, to artificially inflate the output of researchers.

Honorary Authorship and ghost authorship, i.e. listing individuals as authors who do not meet authorship criteria or not naming those who do, are forms of scientific misconduct which undermine the accountability of authorship and authorship as an indicator of scientific productivity. Flanagin et al. (1998) reported that honorary authors, also called guest or gift authors, appeared in 19.3% of a sample of 809

---

[4] http://stadium.open.ac.uk/webcast-ou/documents/20130410_Jonathan_Adams_Presentation.ppt

medical journal articles with US corresponding authors published in 1996, while 11.5% had ghost authors. A more recent study based on 2008 papers showed that the share of papers involving ghost authorship (7.9%) had significantly decreased, while that with honorary authors (17.6%) remained similar (Wislar et al., 2011). Honorary authorship represent one of the most unethical ways to increase publication output, since researchers are added to the author list who have not contributed substantially to the paper. In some extreme cases, authorship was even for sale. Hvistendahl (2013) reports about an academic black market in China, where authorship on papers accepted for publication in Impact Factor journals were offered for as much as US$ 14,800. Some journals try to prevent this unethical practice by publishing statements of author contributions. However, for the sample of papers in Wislar et al. (2011) publishing author contributions did not have a significant affect on misappropriate authorships. Similarly, within the context of international university rankings, universities in Saudi Arabia have been offering "part-time" contracts of more than US$ 70,000 a year to highly-cited researchers, whose task are simply to update their highly-cited profile on Thomson Reuters' with the additional affiliation, as well as sign the institution's name on their papers, with the sole purpose of increasing the institution ranking in the various university rankings (Bhattacharjee, 2011). In a manner similar to salami publishing, these practices inflate scholars' research output and distort the adequacy of the indicator regarding the concept, i.e., the research activity of authors and institutions which it aims to measure.

Self-citations To a certain extent, author self-citations are natural, as researchers usually build on their own previous research. However, in the context of research evalution, where citations are used as a proxy for impact on the scientific community, self-citations are problematic as they do in fact not mirror influence on the work of other researchers and thus distort citation rates (Asknes, 2003; Glänzel et al., 2006). They are also the most common and easiest way to artificially inflate one's scientific impact. Studies found that author self-citations can account for about one quarter to one third of the total number of citations received within the first three years, depending on the field, but generally decrease over time (e.g., Aksnes, 2003; Glänzel et al., 2006; Costas et al., 2010). The common definition of author self-citations considers mutual (co-)authors of citing and the cited papers, i.e. a self-citation occurs if the two sets of authors are not disjoint (Snyder & Bonzi, 1998). Self-citations can be removed from bibliometric studies to prevent distortions particularly on the micro and meso level. The correlations between citations and co-citations on a larger aggregation level were shown to be strong so that it is not necessary to control for self-citations on the country level (Glänzel & Thijs, 2004). Another way to artificially increase one's citation counts, which cannot be easily detected, is citation cartels, where authors agree to cite each others' papers.

Increasing the Journal Impact Factor Due to its importance, the Impact Factor is probably the most misused and manipulated indicator. There are several ways how journal editors "optimize" the Impact Factor of their periodicals, a phenomenon referred to as the 'numbers game' (Rogers, 2002), 'Impact Factor game' (The PLoS Medicine Editors, 2006) or even 'Impact Factor wars' (Favaloro, 2008). One method is to increase the number of citations to papers published in the journal in the last two years, i.e. journal self-citations, by pushing authors during the peer-review process to enlarge their reference lists (Seglen, 1997a; Hemmingsson et al., 2002). Thomson Reuters' monitors journal self-citations and suspends periodicals suspected of gaming. In the 2012 edition of the JCR, 65 titles were red-flagged[5] and thus not given an Impact Factor. Editors of four Brazilian journals went even a step further and formed a citation cartel to inflate each other's Impact Factors through *citation stacking*, which is not as easily detectable as journal self-citations (van Norden, 2013). Another approach to manipulate the indicator is to foster the publication of non-citable items, which collect so-called 'free citations' to the journal by adding to the numerator but not the denominator (Moed & van Leeuwen, 1995; Seglen, 1997a).

---

[5] http://admin-apps.webofknowledge.com/JCR/static_html/notices/notices.htm

Cumulative or Personal Impact Factors Aside from the Impact Factor being a flawed journal indicator, its worst application is that of cumulative or personal Impact Factors. Developed out of the need to obtain impact indicators for recent papers, which have not yet had time to accumulate citations, the journal Impact Factor is used as a proxy for the citations of papers published in the particular journal. The problem with using the journal Impact Factors as an expected citation rate is that due to the underlying skewed distributions, it is neither a predictor nor good representative of actual document citations (Seglen, 1997a; Moed, 2002). Recent research also provided evidence that this predictive power is actually decreasing since the 1990s (Lozano, Larivière and Gingras, 2013). Although the citations of papers determine the journal citation rate, the opposite does not apply. To go back to the Acta Crystallographica A example mentioned in section 3.3.1, of the 122 articles and reviews published in 2007 and 2008, only the one highly cited document (49 citations) obtained as many citations than the Impact Factor value of 2009 (49.926), all other documents were cited less than 17 times, and 47% were not cited at all during that year. Even though this shows that the Impact Factor is not at all a good predictor of citation impact, the cumulative Impact Factor, i.e. adding up Impact Factors for the papers published by a researcher, is frequently applied, most often in the biomedical fields, where grant committees ask cumulative Impact Factors of applicants and researchers list them in their CVs. Despite these deficiencies, the Impact Factor is still applied as a "cheap-and-cheerful" (Adam, 2002, p. 727) surrogate for actual citations because it is available much faster. Although proven to be meaningless (Seglen, 1997b), financial bonuses are awarded and hiring and tenure decisions based on the cumulative Impact Factors (Adam, 2002; Rogers, 2002; The PLoS Medicine editors, 2006; Cameron, 2005). It is hoped that the recent San Francisco Declaration on Research Assessment (DORA) can put an end to this malpractice.

5. Conclusions

This chapter has reviewed the framework, methods and indicators used in bibliometrics, focusing on its application in research evaluation, as well some of its adverse effects on researchers' scholarly communication behavior. It has argued that such indicators should be interpreted with caution, as they do not represent research activity – let alone scientific impact – but, rather, are indicators of such concepts. Also, they are far from representing the whole spectrum of research and scientific activities, as research does not necessarily lead to publication. Along these lines, bibliometric indicators do not provide any insights on the social or economic impact of research and are, thus, limited to assessing the impact of research within the scientific community. Hence, these indicators have to be triangulated and applied carefully, adapted to the units that are assessed. For example, while bibliometric data could be quite meaningful for assessing the research activity of a group of physicists in the United States, it would likely be much less relevant for historians in Germany, who typically publish in books national journals. There is not a one-size-fits-all bibliometric method for research evaluation but, rather, several types of methods indicators that can be applied to different contexts of evaluation and monitoring. On the whole, these indicators can certainly not offer a legitimate shortcut to replace traditional peer review assessments, especially at the level of individual researchers and small research groups. However, to assess the global output on the meso or macro level, bibliometrics can be quite useful, as it is perhaps the only method that can be used to compare and estimate strengths and weaknesses of institutions or countries. Most importantly, entities, such as researchers and institutions, should not be ranked by one indicator, but multiple metrics should be applied to mirror the complexity of scholarly communication. Moreover, quantitative metrics need to be validated and complemented with expert judgements. In addition to the importance of the application context of bibliometrics, attention should also be paid to data quality. This involves using adequate databases such as Web of Science or Scopus – instead of a blackbox like Google Scholar – and data cleaning in particular regarding author names and institutional addresses. Normalized indicators need to be used to balance out field, age and and document type biases.

The Impact Factor has dominated research evaluation far too long[6] because of its availability and simplicity, and the h-index has been popular because of a similar reason: the promise to enable the ranking of scientists using only one number. For policy-makers – and, unfortunately, for researchers as well – it is much easier to count papers than to read them. Similarly, the fact that these indicators are readily available on the web interfaces of the Web of Science and Scopus add legitimacy to them in the eyes of the research community. Hence, in a context where any number beats no number, these indicators have prevailed, even though both of them have long been proven to be flawed. The same could happen to new social-media based metrics, so-called altmetrics (see Weller's chapter in this book). The reason the popularity of indicators such as the Impact Factor and h-index is that alternatives are not as simple and rightly available. Relative indicators that adequately normalize for field and age differences and percentile indicators that account for the skewed citation distributions require access to local versions of the Web of Science, Scopus or other adequate citation indexes, and are much more difficult to understand. Multidimensional approaches are more complex than a simple ranking according to one number. Still, this is the only way fair and accurate evaluations can be performed.

The current use of simple indicators at the micro level – such as the Impact Factor and the h-index – has side effects. As evaluators reduce scientific success to numbers, researchers are changing their publication behavior to optimize these numbers through various unethical tactics. Moreover, the scientific community has been, since the beginning of the twentieth century, independent when it comes to research evaluation, which was performed through peer-review by colleagues who understood the content of the research. We are entering a system where numbers compiled by private firms are increasingly replacing this judgement. And that is the worst side effect of them all: the dispossession of researchers from their own evaluation methods which, in turn, lessens the independence of the scientific community.

References

Abt HA (1992) Publication practices in various sciences. Scientometrics 24(3):441–447. doi: 10.1007/BF02051040

Adam D (2002) The counting house. Nature 415(6873):726–729

Aksnes DW (2003) A macro study of self-citation. Scientometrics 56:235–246

Archambault E, Larivière V (2009) History of the journal Impact Factor: Contingencies and consequences. Scientometrics 79(3):635–649. doi:10.1007/s11192-007-2036-x

Bergstrom CT (2007) Eigenfactor: Measuring the value and prestige of scholarly journals. College & Research Libraries News 68(5):314–316

Bhattacharjee, Y (2011) Saudi Universities Offer Cash in Exchange for Academic Prestige. Science 334(6061): 1344–1345

Biagioli M (2003) Rights or rewards? Changing frameworks of scientific authorship. In: Biagioli M, Galison P (eds) Scientific authorship: Credit and intellectual property in science. Routledge, New York / London, p 253–279

Birnholtz J (2006) What does it mean to be an author? The intersection of credit, contribution and collaboration in science. Journal of the American Society for Information Science and Technology 57(13):1758–1770.

Borgman CL Furner J (2002) Scholarly communication and bibliometrics. Annual Review of Information Science and Technology 36:3–72

Bornmann L, Mutz R (2011) Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. Journal of Informetrics 5(1):228–230. doi: 10.1016/j.joi.2010.10.009

Bradford SC (1934) Sources of information on specific subjects. Engineering 137:85–86. doi: 10.1177/016555158501000407

Burton RE, Kebler RW (1960) The half-life of some scientific and technical literatures. American Documentation, 11(1):18–22. doi:10.1002/asi.5090110105

Calver MC, Bradley JS (2009) Should we use the mean citations per paper to summarise a journal's impact or to rank journals in the same field? Scientometrics 81(3):611–615. doi: 10.1007/s11192-008-2229-y

Cameron BD (2005) Trends in the usage of ISI bibliometric data: uses, abuses, and implications. Portal - Libraries and the Academy 5(1):105–125. doi: 10.1353/pla.2005.0003

Cole JR, Cole S (1967) Scientific output and recognition: a study in the operation of the reward system in science. American Sociological Review 32(3):377–390

---

[6] Only recently the San Francisco Declaration on Research Assessment (DORA) took a stand against the use of the Impact Factor in article and author evaluation.

Cole JR, Cole S (1973) Social Stratification in Science. University of Chicago Press, Chicago

Cole FJ, Eales NB (1917) The history of comparative anatomy. Part I: A statistical analysis of the literature. Science Progress 11(43):578–596

Costas R, van Leeuwen TN, Bordóns M (2010) Self-citations at the meso and individual levels: effects of different calculation methods. Scientometrics 82(3):517–537

Cronin B, Overfelt K (1994) Citation-based auditing of academic performance. Journal of the American Society for Information Science 45(3):61–72

De Bellis N (2009) Bibliometrics and Citation Analysis. From the Science Citation Index to Cybermetrics. The Scarecrow Press, Lanham / Toronto / Plymouth

De Lange C, Glänzel W (1997) Modelling and measuring multilateral coauthorship in international scientific collaboration. Part I. Development of a new model using a series expansion approach. Scientometrics 40(3):593–604.

Engels TCE, Ossenblok TLB, Spruyt EHJ (2012) Changing publication patterns in the Social Sciences and Humanities, 2000–2009 Scientometrics 93(2): 373-390. doi: 10.1007/s11192-012-0680-2

Favaloro EJ (2008) Measuring the quality of journals and journal articles: The Impact Factor tells but a portion of the story. Seminars in Thrombosis and Hemostasis 34(1):7–25. doi: 10.1055/s-2008-1066030

Flanagin A, Carey LA, Fontanarosa PB, Phillips SG, Pace BP, Lundberg GD, Rennie D (1998) Prevalance of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA-Journal of the American Medical Association 280(3):222–224

Garfield E (1964) "Science Citation Index" – A new dimension in indexing. Science 144(3619):649–654

Garfield E (1972) Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies. Science 178(4060):471–479

Garfield E (1979) Citation Indexing - Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons Inc., New York / Chichester / Brisbane / Toronto

Garfield E (1983) Citation Indexing. Its Theory and Application in Science, Technology, and Humanities. ISI Press, Philadelphia

Gingras Y, Larivière V (2011) There are neither "king" nor "crown" in scientometrics: Comments on a supposed "alternative" method of normalization. Journal of Informetrics 5(1):226–227

Glänzel W, Debackere K, Thijs B, Schubert A (2006) A concise review on the role of author self-citations in information science, bibliometrics and science policy. Scientometrics 67:263–277

Glänzel W, Schubert A, Thijs B, Debackere K. (2011) A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. Scientometrics 87(2):415–424. doi: 10.1007/s11192-011-0345-6

Glänzel W, Thjis B (2004) The influence of author self-citations on bibliometric macro indicators. Scientometrics 59(3):281–310

Gonzalez-Pereira B, Guerrero-Bote V P, de Moya-Anegon F (2010) A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics 4(3):379–391. doi: 10.1016/j.joi.2010.03.002

Gross PLK, Gross EM (1927) College libraries and chemcial education. Science 66(1713):385–389. doi: 10.1126/science.66.1713.385

Haustein S (2012) Multidimensional Journal Evaluation. Analyzing Scientific Periodicals beyond the Impact Factor. De Gruyter Saur, Berlin

Hemmingsson A, Mygind T, Skjennald A, Edgren J (2002) Manipulation of Impact Factors by editors of scientific journals. American Journal of Roentgenology 178(3):767–767

Higgs D (2013) One size doesn't fit all: On the co-evolution of national evaluation systems and social science publishing. Confero 1(1):67-90

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102(46):16569–16572. doi: 10.1073/pnas.0507655102

Hvistendahl M (2013) China's publication bazaar. Science 342(6162):1035–1039. doi: 10.1126/science.342.6162.1035

King J (1987) A review of bibliometric and other science indicators and their role in research evaluation 13(5):261–276. doi: 10.1177/016555158701300501

Larivière V, Gingras Y (2010) On the prevalence and scientific impact of duplicate publications in different scientific fields (1980-2007). Journal of Documentation 66(2):179-190.

Larivière V, Gingras Y (2011) Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation. Journal of Informetrics 5(3):392–399. doi: 10.1016/j.joi.2011.02.001

Larivière V, Gingras Y, Archambault E (2006) Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3): 519–533.

Larivière V, Gingras Y, Archambault E (2009) The decline in the concentration of citations, 1900-2007. Journal of the American Society for Information Science and Technology 60(4): 858–862. doi: 10.1002/asi.21011

Leydesdorff L (2009) How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology 60(7):1327–1336. doi: 10.1002/asi.21024

Leydesdorff L, Opthof T (2010) Normalization at the field level: Fractional counting of citations. Journal of Informetrics, 4(4):644–646. doi:10.1016/j.joi.2010.05.003

Lotka AJ (1926) The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16(12):317–323

Lozano, G.A., Larivière, V., Gingras, Y. (2012). The weakening relationship between the Impact Factor and papers' citations in the digital age. Journal of the American Society for Information Science and Technology 63(11): 2140-2145.

Martin BR (2013) Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. Research Policy 42:1005–1014. doi: 10.1016/j.respol.2013.03.011

Martin BR, Irvine J (1983) Assessing basic research. Some partial indicators of scientific progress in radio astronomy. Research Policy 12:61–90

Martyn J, Gilchrist A (1968). An Evaluation of British Scientific Journals. Aslib, London

Merton RK (1973) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago

Merton (1977) The sociology of science: an episodic memoir. In: Merton RK, Gaston J (eds) The Sociology of Science in Europe. Southern Illinois University Press, Carbondale, p 3–141

Merton RK (1988) The Matthew effect in science, II. Cummulative advantage and the symbolism of intellectual property. ISIS, 79:606-623

Moed HF (2005) Citation Analysis in Research Evaluation. Springer, Dordrecht

Moed HF (2009) New developments in the use of citation analysis in research evaluation. Archivum Immunologiae et Therapiae Experimentalis 57(1):13–18. doi: 10.1007/s00005-009-0001-5

Moed HF (2010) Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4(3):265–277. doi: 10.1016/j.joi.2010.01.002

Moed HF, van Leeuwen TN (1995) Improving the accuracy of Institute for Scientific Information's journal Impact Factors. Journal of the American Society for Information Science, 46(6):461–467

Moed HF, van Leeuwen TN, Reedijk J (1999) Towards appropriate indicators of journal impact. Scientometrics 46(3):575–589. doi: 10.1007/BF02459613

Norris M, Oppenheim, C (2003) Citation counts and the Research Assessment Exercise V: archaeology and the 2001 Research Assessment Exercise. Journal of Documentation 59(6):709–730.

Otlet P (1934) Traité de documentation: le livre sur le livre, théorie et pratique. Editiones Mundaneum, Brussels

Pontille D (2004) La Signature Scientifique: Une Sociologie Pragmatique de l'Attribution. CNRS Éditions, Paris

Price, DJDS (1963) Little Science, Big Science. Columbia University Press, New York

Price, DDS (1981) Letter to the editor: Multiple authorship. Science 212(4498):984–1066

Pritchard A (1969) Statistical Bibliography or Bibliometrics? Journal of Documentation 25(4): 348–349

Rinia EJ, van Leeuwen TN, van Vuren HG, van Raan AFJ (1998) Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. Research Policy 27:95–107

Rogers LF (2002) Impact Factor: The numbers game. American Journal of Roentgenology 178(3):541–542. doi: 10.2214/ajr.178.3.1780541

Rowlands I, Nicholas D (2005) New Journal Publishing Models – An International Survey of Senior Researchers, CIBER, University College London, London. Available from: http://www.homepages.ucl.ac.uk/~uczciro/pa_stm_final_report.pdf,

Schubert A, Braun T (1996) Cross-field normalization of scientometric indicators. Scientometrics 36(3):311–324

Seglen PO (1992) The skewness of science. Journal of the American Society for Information Science 43(9):628–638

Seglen, PO (1997a) Citations and journal Impact Factors: questionable indicators of research quality. Allergy 52(11):1050–1056

Seglen, PO (1997b) Why the impact factor of journals should not be used for evaluating research. British Mecial Journal 314(7079): 498–502. doi:10.1136/bmj.314.7079.497

Snyder H, Bonzi S (1998) Patterns of self-citation across disciplines. Journal of Information Science 24(6): 431–435

Small H (1987) The significance of bibliographic references. Scientometrics 8:321–340

Smalheiser NR, Torvik VI (2009) Author name disambiguation. Annual Review of Information Science and Technology 43:1–43. doi:10.1002/aris.2009.1440430113

The PLoS Medicine editors. (2006). The Impact Factor game - It is time to find a better way to assess the scientific literature. PLoS Medicine 3(6):e291. doi:10.1371/journal.pmed.0030291

Thomas PR, Watkins DS (1998) Institutional research rankings via bibliometric analysis and direct peer review: A comparative case study with policy implications. Scientometrics 41(3):335–355. doi: 10.1007/BF02459050

Tijssen RJW, Visser MS, van Leeuwen TN (2002) Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? Scientometrics 54(3):381–397

Todorov R, Glänzel W (1988) Journal citation measures - a concise review. Journal of Information Science 14(1):47–56. doi: 10.1177/016555158801400106

Van Norden R (2013) Brazilian citation scheme outed. Thomson Reuters suspends journals from its rankings for 'citation stacking'. Nature 500(7464):510–511

Waltman L, van Eck NJ (2010) The relation between eigenfactor, audience factor, and influence weight. Journal of the American Society for Information Science and Technology 61(7):1476–1486. doi: 10.1002/asi.21354

Waltman L, van Eck NJ (2012) The inconsistency of the h-index. Journal of the American Society for Information & technology 63(2):406–415. doi: 10.1002/asi.21678

Weingart P (2005) Impact of bibliometrics upon the science system: Inadvertent consequences? Scientometrics 62(1):117–131. doi: 10.1007/s11192-005-0007-7

White HD, McCain KW (1989) Bibliometrics. Annual Review of Information Science and Technology 24:119–186

Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD (2011) Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ 343:6128. doi: 10.1136/bmj.d6128

Zipf GK (1949) Human Behaviour and the Principle of Least Effort. Addison-Wesley Press, Reading

Zuckerman, H (1987) Citation analysis and the complex problem of intellectual influence. Scientometrics 12:329–338

Zitt M (2010) Citing-side normalization of journal impact: A robust variant of the Audience Factor. Journal of Informetrics 4(3):392–406. doi: 10.1016/j.joi.2010.03.004