6.9: Bibliometrics for journal editors – an overview

Jenny Neophytou
Bibliometrics Analyst, Wiley, Oxford, UK; jneophytou@wiley.com

Introduction
The past 10 years have seen a proliferation in the number and type of metrics available to measure academic research. The most widely used of these is still the impact factor, which was first mentioned by Eugene Garfield in 1955. The impact factor is one of a range of citation metrics that are based on the premise that, when one author cites an article, that article has had an ‘impact’. It is often used as a way to estimate a journal’s influence on its subject area – and, by proxy, its perceived quality.

The use of citation metrics has increased substantially in recent years. Although initially intended as a way to help academics ‘filter content’, they are frequently used to measure academic achievement at both an individual and an institutional level. Many national funding decisions are supported by citation and publication metrics. (Examples include the national Norwegian model for research performance, the upcoming UK Research Excellence Framework and the Chinese academic evaluation system.) This situation has drawn significant criticism, as such practices often encourage conformity in research, ‘salami-slicing’ (or incremental publication) to increase publication output, and the unequal treatment of academics from naturally low-cited disciplines. This means that, despite the influence they exert, citation metrics remain widely misinterpreted and misused. Failure to set appropriate controls, lack of understanding, and deliberate manipulation have all led to the use of misleading data that directly impact on the lives and activities of researchers, with the unintended consequence that bibliometrics have in some cases dictated, rather than reflected, the research environment.

Principles of citation indexing
Citation values can be obtained from a number of services, including Google Scholar, Microsoft Academic Search, CrossRef, PubMed Central and Altmetrics; however, the majority of journal analysis is based on multidisciplinary indexing databases, such as Web of Science and Scopus. The main reason for this is one of control. Unlike autonomous databases such as Google Scholar, content is indexed in the Web of Science and Scopus only after it has been reviewed for academic quality. As such, we know that a citation in one of these databases is derived from academic material. Moreover, since these autonomous databases do not usually provide a list of what they include, we are unable to measure the relative value of a citation. If a database does not index our main competitor journals, we need to be aware that it will not index a large proportion of our citations; however, if we do not know the coverage, we have no way of estimating what proportion of our citations are indexed. This becomes clear when we consider how citations are counted.

When an article is indexed, the database records not only the article metadata but also the article reference list. The presence of the reference list allows the database to create links between cited and citing articles. When these links are formed, a citation is recorded. Citations will therefore be counted in a database only if both the cited and the citing article are indexed, meaning that citation scores are likely to be higher within larger databases. For this reason, the same article is likely to have a larger citation count in Google Scholar than in either Scopus or Web of Science, not because the database has ‘missed’ citations but because it does not index all the citing content.

Citation metrics – necessary controls
At their simplest, most citation metrics are a measure of the average number of citations per paper in a given set of articles. Many factors can, however, influence the rate of citation of any given piece of research. These include the age of the article, the type of research conducted, and the subject area, all of which must be controlled to allow meaningful comparisons between journals (or academics). Application of these controls is evolving constantly, giving rise to new metrics that allow us to look at different aspects of citation and publication behaviour.

The citation window
Many editors request a list of ‘top-cited articles’ to help them evaluate the impact and content of the material published in their journals. Comparing papers by the total number of citations they have received is, however, of limited value, as any measure calculated from total citation counts will be heavily biased towards older papers, not because the research is more ‘citable’, but because it has been around for a longer time. Most metrics set a ‘citation window’, that is, a set period of time after an article’s publication, during which citations will be included in the calculation for that metric. The impact factor, for example, has a 2-year citation window. For an article published in 2010, only citations received in the 2 years after publication (2011 and 2012) will count towards an impact factor numerator (each impact factor being calculated on the basis of the citations in the current year to items published in the 2 previous years). This means that the age of an article is to some extent controlled. As the metric works for the calendar year (or cover year), a paper published in January 2010 has an advantage over a paper published in December 2010.

The appropriate length of the citation window is a matter of constant debate, due largely to differences in citation and publication behaviour between (and within) subjects. Research published within a rapidly evolving field such
as applied physics will probably receive citations far more quickly than research in linguistics. It is unfair to compare the impact factors of two journals in different disciplines.

Figure 1 highlights the difficulties, showing the distribution of 2011 citations by the publication year of the cited article. While in ‘cell biology’ the majority of 2011 citations were to articles falling in the impact factor window, content cited in the ‘geology’ and ‘hospitality, leisure, sport and tourism’ categories tends to be older. This means that the impact factor is measuring a different proportion of a journal’s total number of citations in each discipline.

In an attempt to compensate for this disparity, Thomson Reuters launched the 5-year impact factor in the 2007 Journal Citation Reports, with a 5-year citation window instead of the former 2-year citation window. This makes more likely that a discipline will receive the majority of its citations before the window expires. However, differences in citation and publication culture, not to mention differences in subject coverage, mean that it is unwise to compare most citation metrics across disciplines (but see the alternative metrics section below).

**Document types**

Different types of document tend to be cited at different rates. This is not always a comment on the quality of the research published but merely the nature of the research. However useful they are to the academic community, case reports attract few citations; they are of practical use, but rarely of use in published research. In contrast, review papers tend to receive the most citations; it is, after all, easier to read and cite a single paper that summarizes a whole area than to cite the 50 papers it summarizes. For this reason, many top-ranked journals are actually review journals, so that their impact factor is as much a consequence of their scope as it is of their ‘quality’.

To provide at least some level of control, Thomson Reuters and Scopus both distinguish ‘citable content’ from ‘non-citable content’. While every item indexed in the databases is known as a ‘source item’, only certain document types are defined as ‘citable’ (expected to receive citations). These document types are defined as articles, proceedings papers (or in the case of Scopus, conference proceedings), and reviews. Other content (such as editorials, notes, and meeting abstracts) are classed as ‘non-citable’ and do not count towards the denominator for most metrics. However, there is as yet no control offered for different types of ‘citable’ content, meaning that review journals have a significant advantage in terms of citation metrics. Document types in the Web of Science and Scopus are not necessarily the same as those in the journal. They are allocated on the basis of criteria including:

- the length of the reference list;
- the journal section;
- keywords in the title or abstract field;
- the presence or content of the abstract.

For example, documents in the Web of Science are generally classified as ‘review’ papers if they contain >100 references, or contain key words that suggest a review structure. Similarly, documents are generally classified as ‘proceedings papers’ if the author indicates that a portion of the paper was previously presented at a conference. Of course, as Harzing points out, this process is far from perfect. Mistakes do happen, leading to further distortion between individual journals and disciplines, and data correction requests are regularly submitted.

**Self-citation**

Self-citation can occur at both individual and journal levels. In the context of the impact factor, self-citation refers to the process by which articles in a given journal cite other research published in the same journal. It is usually inevitable. Natural factors influencing the self-citation rate include:

- journal discipline;
- journal type (i.e. niche or archive publications);
- discipline coverage in the citation database.

For example, when a journal in a discipline that has low coverage is indexed, it is likely that many of its competitor titles are not included in the database. This means that many citations to that journal will not be recorded and, consequently, that the majority of visible citations will be self-citations.

Self-citation becomes a problem when it is deliberately inflated in an attempt to improve citation metrics. Such practices distort the citation pool, artificially increasing a journal’s impact factor and leading to misrepresentation in the Journal Citation Reports. To combat this problem, Thomson Reuters publish the self-citation rate of each journal in the Journal Citation Reports and an impact factor with all self-citations removed. Where self-citation is considered to have an undue influence on a journals rank, Thomson Reuters reserves the right to suppress that title for a minimum of 2 years.

Other citation metrics can control, at least to a certain extent, self-citation. Scientific Journal Rankings, for example, limits self-citation to 33% for each journal (other self-citations being removed from the calculation).

**Other citation metrics**

As mentioned previously, there are many different citation metrics, and each sets out to measure a different aspect of citation and publication behaviour (see Table 1). Depending on their goal, these metrics apply differing controls for factors such as the citation window, the publication window, document type, and subject area. The Eigenfactor, for example, is designed to measure the citation network. It uses a 5-year citation window to account for differences in behaviour between disciplines and (as also mentioned above) excludes self-citations, as they measure the effect on the self rather than on the network. In addition, the Eigenfactor uses weighted citations to reflect differences in the citation network. Metrics weights by two main methods: by journal impact and by subject.

**Weighting metrics by citing journal impact**

One problem with citation metrics is that, in most cases, all citations are treated equally. However, it has been argued that citations of high-impact journals should be considered more valuable in any citation metric. The Eigenfactor and article influence score (calculated from Journal Citation Reports data) both weight citation counts according to the impact of the citing journal, that is they assess the ‘impact’ of the citation network rather than just raw citation counts.

Thus, a citation in a high-ranking journal (such as Nature) will score higher than one in a lower-ranking journal, so that the score of any journal is determined not only by the volume of citations but also where they occur.

**Table 1: Citation metrics available**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Data source</th>
<th>Citation window</th>
<th>Weighted citations</th>
<th>Self-citations included</th>
<th>Document types included in denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact factor</td>
<td>Web of Science</td>
<td>2 years</td>
<td>No</td>
<td>Yes</td>
<td>Articles, proceedings, reviews</td>
</tr>
<tr>
<td>5-year impact factor</td>
<td>Web of Science</td>
<td>5 years</td>
<td>No</td>
<td>Yes</td>
<td>Articles, proceedings, reviews</td>
</tr>
<tr>
<td>Immediacy index</td>
<td>Web of Science</td>
<td>0 years (year of publication)</td>
<td>No</td>
<td>Yes</td>
<td>Articles, proceedings, reviews</td>
</tr>
<tr>
<td>Eigenfactor</td>
<td>Web of Science</td>
<td>5 years</td>
<td>Yes – by citing journal quality</td>
<td>No</td>
<td>Articles, proceedings, reviews</td>
</tr>
<tr>
<td>Article influence</td>
<td>Web of Science</td>
<td>5 years</td>
<td>Yes – by citing journal quality</td>
<td>No</td>
<td>Articles, proceedings, reviews</td>
</tr>
<tr>
<td>Source normalized impact per paper (SNIP)</td>
<td>Scopus</td>
<td>5 years</td>
<td>Yes – by subject</td>
<td>Yes</td>
<td>Articles, conference papers, reviews</td>
</tr>
<tr>
<td>Scientific journal rankings (SciMago journal rank)</td>
<td>Scopus</td>
<td>3 years</td>
<td>Yes – by subject</td>
<td>35% maximum</td>
<td>Articles, conference papers, reviews</td>
</tr>
<tr>
<td>H-Index</td>
<td>Not defined</td>
<td>None specified</td>
<td>No</td>
<td>Yes</td>
<td>Not defined</td>
</tr>
</tbody>
</table>
6.9: Bibliometrics for journal editors – an overview

How to get indexed
In most cases, application to any indexing service should be facilitated by the journal’s publisher (where appropriate), to ensure that the correct communications, permissions, and systems are in place in the event of acceptance. Many publishers can provide feedback on your journal that will allow you to estimate the likelihood of acceptance. For example, Thomson Reuters reviews a range of factors before deciding whether to index it in the journal. It is important to ensure that these criteria are met before submitting a journal for coverage. The criteria are:

- Timeliness of publication: Late or short publication can indicate poor academic reception. This means in some journals, the journal will fail in the near future;
- Peer review: The journal must have a robust peer review system in order to maintain research quality;
- Subject area: The journal must be unique. Thomson Reuters must consider that indexing the title will enrich its database rather than being a superfluous addition;
- Internationally: Unless it declares itself to be a regional title, the journal’s author base and editorial board should reflect the geographical diversity of the subject area;
- Citations: Journals are often rejected because of low citation levels for their designated categories. This may simply be due to the fact that it is largely uncited; however, it may also be that the journal’s main competitors are not indexed, so that Thompson Reuters has no record of articles that cite the journal.

Usage metrics: COUNTER
Usage metrics have become increasingly popular, leading to much discussion over use of a ‘usage factor’ rather than the ‘impact factor’. The proposed metric is similar to the impact factor. It calculates median usage in a census period (such as 2010–2011) of items published in a specified year range (2010–2011). Just as different citation databases report on different citation numbers, so can different usage systems report different usage. For this reason, Counting Online Usage (COUNTER) has established a range of standards and codes of practice for organizations reporting usage data. Only platforms that are ‘COUNTER-compliant’ will be eligible to submit data for calculation of the usage factor; in theory, all usage data so used will be both comparable and compatible. Many factors should be considered in looking at usage data. Just as with citation metrics, it is important to be aware of the factors that can influence the rate of download:

- Defining usage: When should an online event be measured as ‘usage’?
- Controlling for robots and web-crawlers: Much usage is not genuine academic accessing research but by automated robots and data-miners. These also appear as online events, and can artificially inflate usage. Robust systems must be in place to exclude them from usage reports;
- Data sources: Some journals are available on multiple platforms, not all of which are COUNTER-compliant. Will such journals have an unfair disadvantage?

• ‘Self-usage’ and promotional usage: Just as self-citation can distort citation metrics, so can self-usage and promotional usage affect usage metrics. How can we control for these practices?
• Geographical disparity: One problem with any type of online-only metric is variation in internet access across the world. Journals and articles that target geographical locations with poor levels of internet access would therefore be disadvantaged, just as many ‘altmetrics’ prioritize regions with widespread access to social media.

Even after these factors are accounted for, we are left with issues of the definition of ‘usage’. What counts as usage and what counts as citation? What factors influence the rate of usage? How can we mitigate the influence of these factors?

Conclusion
When using metrics to analyse a paper, publication, academic, or institution, it is essential that the full range of influence be considered. Citation metrics (particularly the impact factor) have drawn significant criticism, leading to the development of metrics from new data sources. However, none of these sources provides a direct alternative to the impact factor, which remains a powerful tool. Usage metrics can provide new insights into different aspects of researcher behaviour. Through such metrics, we are able to observe patterns in research output and uptake, allowing us to monitor trends and practices in the academic market. Bibliometrics can be powerful tools, allowing editors and publishers to monitor trends; however, they can only provide insight, not answers. Used without appropriate awareness, metrics can distort trends in academic research, yielding (and disciplines) that chase metrics yet fail to serve the academic community.

References

Weighting metrics by subject
The ‘source-normalized impact per paper’ (SNIP) is calculated from Elsevier’s Scopus database and is weighted to account for differences in citation behaviour between disciplines. It is introduced to ensure that the metrics for journals in different disciplines can be compared. For example, the citation behaviour of researchers in the biomedical sciences (high rates of citations) is quite different from that of researchers in geology (fewer citations). Therefore, the impact factors of life science journals tend to be higher than those of geological journals, making it difficult to compare the relative ‘success’ in terms of citations between these disciplines. SNIP takes the average citation within each discipline and applies it to normalize the metrics and thus produce an ‘impact factor’ that can be compared with that of any other journal, whether in the same discipline or not.

Weighting metrics by life-long citation
Another citation metric, the H-index, was designed to measure publication and citation behaviour at the author level rather than the journal level (although it is still often used as a journal-level metric!). Its premise is that an academic has an H-index of y if he or she has published γ papers with a minimum of γ citations each; therefore, the metric is effectively limited by the number of papers an academic has produced. An academic who has five papers each with five citations will have the same H-index as one whose five papers each have been cited five times. The traditional H-index sets neither a time-frame for the publications nor a citation window, leading to criticism that the metric is limited by the number of papers that have been in their fields for a longer time. In addition, the lack of a defined citation source (such as Web of Science, Scopus or Google Scholar) means that academics can calculate different values for the H-index for the same set of articles. Since its creation in 2006, many variations on the H-index have been suggested. One, known as the H5-index, is used by Google Scholar in creating their journal-level metrics. While the H5-index still fails to set a citation window (or ‘time period’) for use, it controls for these practices? How can we mitigate the influence of these factors?

Conclusion
When using metrics to analyse a paper, publication, academic, or institution, it is essential that the full range of influencing variables be considered. Citation metrics (particularly the impact factor) have drawn significant criticism, leading to the development of metrics from new data sources. However, none of these sources provides a direct alternative to the impact factor, which remains a powerful tool. Usage metrics can provide new insights into different aspects of researcher behaviour. Through such metrics, we are able to observe patterns in research output and uptake, allowing us to monitor trends and practices in the academic market. Bibliometrics can be powerful tools, allowing editors and publishers to monitor trends; however, they can only provide insight, not answers. Used without appropriate awareness, metrics can distort trends in academic research, yielding (and disciplines) that chase metrics yet fail to serve the academic community.