

# Methodological points of view

- Citation time window
- Fractionalisation
- Normalisation
- The indicators' sensitivity to errors
- Applicability of data for bibliometric analysis

## Citation time window

By 'citation time window' (often shortened to 'citation window') we are referring to the time period analysed, during which the citations were made. Apart from exceptional cases, there is usually no minimum limit for the time window, that is, citations made even before the nominal publication date will be included. For the upper limit, however, a minimum requirement should be specified. Scientific publishing including peer review is a time-consuming process, so the accumulation of citations for articles is slow. It is therefore futile to make impact assessments based on the number of citations for very recent publications. In addition to this, the accumulation rate of citations varies considerably between fields of science. The choice of citation window is also influenced by the intended use: For funding decisions and recruitment, it is not possible to wait several years to find out the final citation impact of a publication, instead, decisions must be made with the best information available at the time. Finally, it should be noted that the sufficient length of the citation window also depends on the impact indicators used.

Typically, the shortest time window used is three years. This means that after the year of publication, there will then be another two years before impact assessments of the publication are carried out. A time window of three years is sufficient for a preliminary assessment in fields where citations are accumulating rapidly. In fields where the publication process is streamlined and citations accumulate very quickly (e.g. some areas of life sciences and medical science), a shorter citation window may be considered if there is frequently updated citation data. In contrast, for fields with a slower publication cycle (many fields of humanities, mathematics), a period of even three years is clearly too short, especially if Top x% indicators are to be used. In these fields, there are situations where, two years after its publication, the publication still reaches the Top 10% of citations with a single citation, which can hardly be interpreted as a sign of citation impact.

The older the publications are at the time of review, the longer they have had to accumulate citations. Naturally, the largest amount of data and the most accurate analysis is obtained by using an open-ended citation window. In this case, all citations to the publication, even if they come decades after its publication, influence the outcome. In bibliometric impact analyses, an open-ended citation window should be used as a priority. For example, limiting citations to only three years after publication, regardless of the age of the publications, will skew the results more significantly the older the publications are. Publications that are quickly receiving citations upon publication (i.e. "hot papers") may not be among the most cited in their control group in a few years' time. The short citation time window therefore measures a different scientific citation impact than the open-ended time window. The short citation window highlights publications that are part of the current scientific debate at the time of their publication, while publications that leave a more lasting impression often only gain attention over a longer period of time.

Furthermore, the reduced citation window does not treat publications from different countries identically: a short time window favours some countries (Germany, the United Kingdom, the United States) and makes the results of other countries (e.g. Finland, Sweden, Norway) look less favourable than they actually are. Therefore, the reduced time window is particularly disadvantageous in studies of the temporal development of the countries' citation impact.

An artificially reduced citation window cannot be justified on the grounds that an unlimited time window would favour older publications. A responsibly conducted analysis will never directly compare the number of citations between publications of different ages. A reduced citation window may be used if the purpose of the analysis is to explicitly identify publications that were considered important at the time of their publication.

## Fractionalisation

Scientific publications are rarely written by one researcher. Researchers from several organisations or countries can be involved in the process. When assessing the **extent of the scientific publishing activities** of an entity (country, university), the result can be presented either as a full count (or whole count), or as fractionalised publication counts. A full publication count answers the question 'How many publications has the organisation/group/author been involved in?', whereas a fractionalised publication count represents the proportion the organisation/group/author accounts for of the total number of publications in the dataset. Fractionalising is most naturally done between units of the same level: when comparing countries, publications are fractionalised between countries, and when comparing organisations, they are fractionalised between organisations. Fractionalised publication counts can be compared, for example, to the publication counts in the entire dataset by field or, in the case of organisations, the fractionalised publication count of the host country. Full publication counts are not comparable in the same way because co-publications may be counted multiple times. For full publication counts, careful consideration must always be given to whether the results are summable at all. Full publication counts can be added up if each publication belongs to no more than one summed subset (e.g. publications released in different years). In contrast, a university's publication count cannot be counted by summing up the university's full publication counts from all fields of science if one publication may be classified in more than one field.

However, when assessing the **citation impact of publishing activity**, the publications and publication-specific impact values must always be fractionalised, i.e. weighted by factors whose sum equals 1.0 for each publication. Unless there are particularly compelling reasons to do otherwise, fractionalisation should be carried out equally with all parties involved in creating the publication. Researchers or research groups may have different roles in different projects (experimental work, field work, theoretical work), and there are no objective indicators to rank them against each other. Seemingly simple options, such as counting author affiliations, are usually artificial and easily manipulated. If impact calculations are made without fractionalisation, the results lack scale (the results no longer indicate how the impact of the examined author entities relates to the average of the whole dataset) and significantly favour the countries, organisations and fields of science that appear most often in the co-publications.

## Normalisation

In publication metrics, normalisation refers to counting publication-specific citation impact values using methods that ensure that the average of the results, both for the whole dataset and for each year and field of science separately, is some easily comparable standard value, usually 1.0. This makes it easier to interpret results. If the unit in question (country, organisation, group) achieves a result above this standard value, it can be interpreted that the entity has succeeded in producing more impactful research than the field average. Without normalisation, the averages of impact indicators vary by field and year, making it difficult to interpret the results.

Normalisation is carried out for publication sets by field and year, sometimes even by publication type. When determining the citation impact of a single publication, it should not at any stage be compared with publications from other fields of science or from different years, otherwise the benefits of normalisation (fixed field-independent benchmark) will be lost. As a consequence, it is not possible, for example, to count a version of the Top 10% indicator that is equitable across fields of science by searching for publications in the Top 10% of citations from the joint distribution of field-normalised citation counts for all fields of science. Instead, the most cited decile must be found separately each year in each field of science, in which case, of course, there is no need to process the citation counts in any way – the order of publications does not change if all citations are divided by the same number.

## The indicators' sensitivity to errors

The data used for bibliometric analyses can generally be regarded as inherently incomplete and inaccurate in some way, except perhaps for very small analyses. It is therefore not at all unimportant how sensitive the indicators used are to errors in the source data. We can ask how big a change removing or adding one publication or citation makes to the outcome, and through this try to assess the reliability of the results. In general, different datasets contain partly different publications and the 'off-the-shelf' pieces of software used to calculate citation impact indicators use different algorithms, so the results also vary depending on the data and the software. The citation time window also affects how much relative change a single citation can produce. A short citation window is clearly more sensitive. The higher number of citations over a longer time window means that the relative importance of a single citation is often relatively small.

**Top x% indicators** are not sensitive to a single, highly cited publication in the target group being reviewed. For example, it is irrelevant to the Top 10% indicator whether the publication is in the top 0.1% or the top 5% most cited in its class. On the other hand, near the limit of the Top 10% class, a single citation is enough to substantially increase the value of the publication. In such cases, the data used can determine the impact value of a publication in one direction or another. The effect may not be relatively significant if the set of publications being analysed is large (country-level analyses), but for small sets of publications (a research group, an individual researcher) the difference can become significant. The same applies, of course, if a publication is missing from the target group in the first place. Since the Top x% indicators are built on the idea that only a small proportion of publications are considered to have any value at all in terms of impact, it is very important to identify all publications by the analysed target group from the data used.

The situation is the opposite for field-normalised citation indicators. Since each citation always adds the same slight amount to the impact value of the publication, the absence of a single citation has little effect on the outcome. On the other hand, a single publication in the target group that has received a significant number of citations in its own field of science can skew the results to the point of being unusable, as there is no upper limit on the impact value of a single publication. In extreme cases, a single publication can multiply the results of an entire organisation.

## Applicability of data for bibliometric analysis

The publication database used for bibliometric analysis must be comprehensive in two ways. Firstly, the data must provide a representative sample of the publications of the unit being reviewed. Therefore, we cannot analyse a business school based on publication data that is solely focused on medical science. The sufficiency of the sample can be evaluated by comparing the number of publications entered in the database with the number of scientific publications found in the unit's own register (**external coverage**). In general, for Finnish universities and research institutes, the external coverage of the most common commercial publication databases is sufficiently high in natural sciences and technology (excluding computer sciences), medical science and agriculture and forestry. For social sciences and humanities, the situation is generally worse, so the results of bibliometric analyses in these fields of science should be interpreted with caution. In addition to the external coverage requirement, which measures relative representativity, an absolute minimum limit of 50 publications is usually applied in the calculation of citation impact indicators. For publication volumes smaller than this, no impact results should be presented at all.

On the other hand, the data must also provide enough comparable data to allow a meaningful evaluation of the citation impact of the unit being reviewed in relation to other publications. We cannot make claims about the impact of publications, even if they happen to be in the database, if most of the publications in the same field of science are excluded from the database. Naturally, it is not possible to directly calculate the proportion of these excluded publications, but it can be estimated using internal coverage. **Internal coverage** indicates the proportion of citations entered in the publication database that are attributed to other publications in the same database. If the internal data coverage is low in a field of science (e.g. less than 40%), a significant proportion of the key publications in that field will most likely be missing from the data used and there will be no justification for impact calculations..