Vetenskapsrådet

Reference no.
113-2014-7357
Date
2015-03-04
Administrators
Elisabeth Sjöstedt
Henrik Aldberg
Carl Jacobsson

Guidelines for using bibliometrics at the Swedish Research Council

Abstract

Bibliometrics is used at the Swedish Research Council to study production volume, publication patterns and scientific impact. The Research Council believes that bibliometric indicators are especially useful within natural sciences and medicine, where current bibliometric databases have good coverage.

The Research Council's bibliometric analyses are generally based on fractionalised articles, where the sum of all fractions corresponds to the actual number of published articles. Without fractionalisation, an article with several authors may for example be counted several times, resulting in an overestimation of both the total scientific production and the impact.

For citation-based indicators, the Research Council uses mean normalised citations. Field normalisation allows you to identify frequently cited articles in all fields, regardless of citation traditions. Self-citation is not included in the calculation of the mean normalised citation rate.

As with all statistics, bibliometrics must be based on a sufficiently large sample size. For pure bibliometric comparisons, the publication data should exceed at least 50 articles, while results which are to be used by subject experts along with other information can be used if the sample size is more than 20 articles.

The Research Council is very restrictive when it comes to bibliometric comparisons between individuals and it does not use the h-index or similar indicators.

In cases where bibliometric data is included in the Council's evaluation of applications, this is only used by experts in the relevant field as part of a larger body of data. Bibliometrics never forms the sole basis for an assessment of the applicants' merits.

Introduction. An important step in all research is to disseminate research results so that others can confirm, apply and build on them. The number of scientific publications in the world has steadily increased over the past 50 years [5]. There is also an increasing amount of resources spent on research, meaning that there is an increasing need for follow-up and studies of the growing research production.

Today, a significant amount of the international scientific article production is registered in databases, which often includes information on who has cited what articles. Bibliometrics is about producing statistics from this information in order to, for example, describe productivity or impact in the form of how often results are cited by others. At the same time, there are research areas which are not covered particularly well by current international bibliometric databases and, in order to describe this research production, bibliometrics as a field needs to be developed further. Other interesting areas of development include how bibliometrics can be used to catch interdisciplinary research, and how ground-breaking work and new fields of research can be identified at an early stage.

This document describes how bibliometrics is used by the Research Council. It also highlights some important limitations of bibliometric methods. It then explains concepts such as fractionalisation, field normalisation and the importance of sufficient publication sample sizes. It closes by describing how the Research Council uses bibliometrics when evaluating applications.

Areas of application. Bibliometrics is used by the Research Council to study production volume, publication patterns and scientific impact. Bibliometrics also allows us to study how these factors change over time. The Research Council believes that bibliometric indicators are especially useful within the natural sciences and medicine, where current bibliometric databases have good coverage, but also in technical sciences and to some degree in psychology and economics.

Peer review can be said to be the standard tool for assessing research quality. In later years, bibliometrics has become increasingly prominent, especially when it comes to comparing extensive research productions, for example between countries or large higher education institutions. In such cases peer review quickly becomes very costly and time consuming. Bibliometrics should on the other hand be used very carefully when studying smaller publication volumes. It is in other words important to know the strengths and weaknesses of both methods, to know when one or the other should be applied and how they complement each other.

When the Research Council assesses research groups or research environments, it often uses expert panels which in turn may use bibliometrics as part of the data when making their assessment. Figure 1 illustrates a rule of thumb.

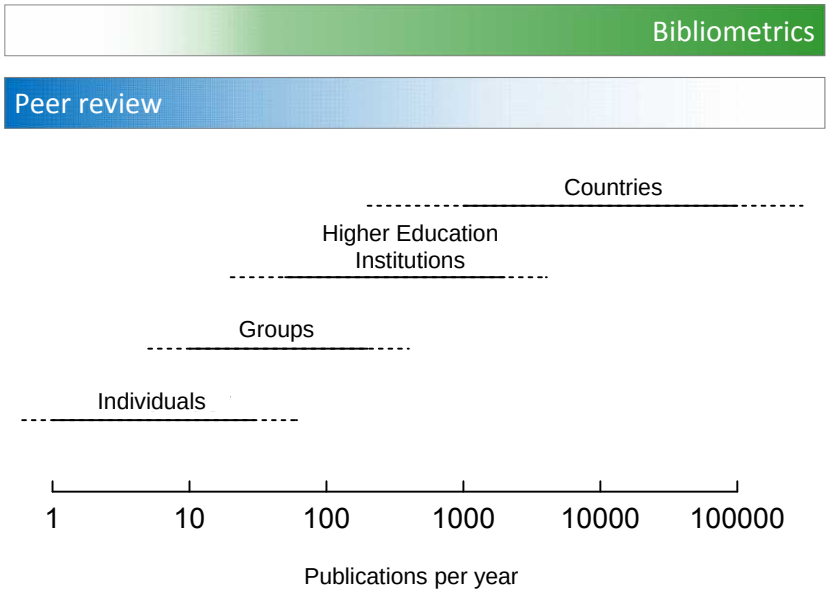


Figure 1: Bibliometrics and peer review have partially different applications when it comes to describing and assessing research production.

Limitations. A bibliometric analysis is based on access to good basic data. In research areas which are not covered well by bibliometric databases, bibliometrics must be used cautiously. The Research Council has an international publication database whose contents approximately correspond to the contents of *Web of Science*.[1,2] Figure 2 shows how well Norwegian research production in various subjects was covered by *Web of Science* in the mid-00s. We can see that the humanities and large parts of the social sciences are largely missing.[3] The coverage will improve to some degree over time as *Web of Science* expands to cover more journals, and to some degree also books, in the field of humanities and social sciences. However, as the large commercial databases mainly focus on English journals, the research production that is published in Swedish channels, and largely in book form, will still be missing in the future. This means that the international databases need to be supplemented by national databases. In Sweden there is SwePub, which collects publication information from a large number of Swedish higher education institutions.[4] The National Library of Sweden, which manages SwePub, has been commissioned by the Government to improve its data quality so that it can be used for bibliometric analyses. The Research Council cooperates with the National Library on this task. As a national database does not include international citations, there is also a need for new bibliometric analysis methods adapted to SwePub's data content.

[1] The Research Council's database is based on the following products: Science Citation Index Expanded, Social Science Citation Index, and Arts and Humanities Citation Index (©Copyright Thomson Reuters).

[2] Two other international databases with broad ambitions are Elseviers Scopus and Google Scholar.

[3] One important reason for the poor coverage of humanities and social sciences is that the international publication databases almost exclusively index scientific journals while the humanities and social sciences often place greater prestige in publishing books. See Appendix B in [7] for more information on the difficulties relating to bibliometrics for the humanities and social sciences.

[4] http://swepub.kb.se/.

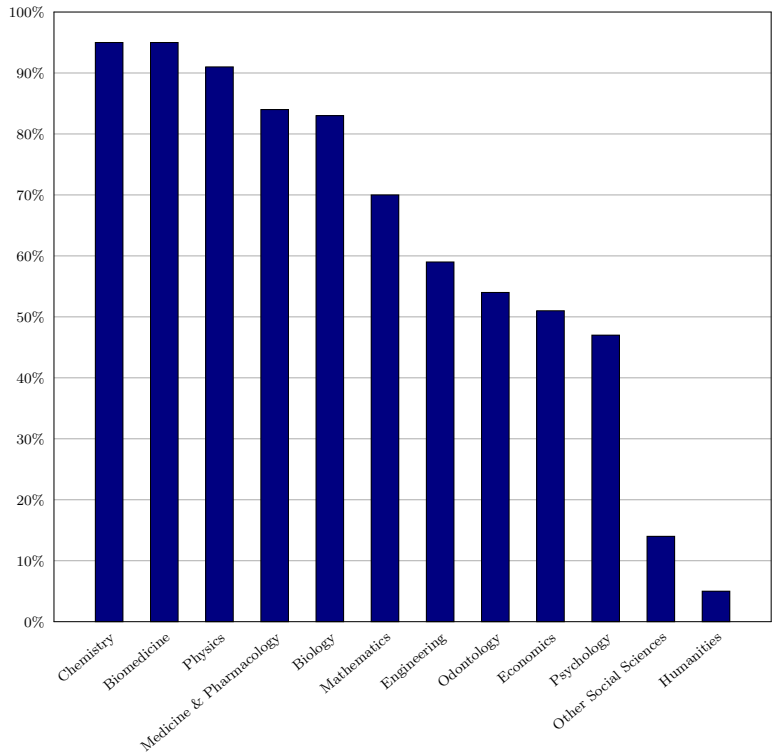


Figure 2: Coverage of Norwegian research production in the international bibliometrics database *Web of Science*.

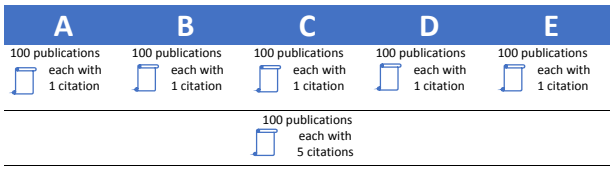
Common indicators. The most common bibliometric indicators used by the Research Council are the number of publications, which is used to estimate research productivity, the number of citations, which helps measure impact, and the mean citation rate, which helps compare impact between, for example, different countries. Another indicator of research impact is looking at the proportion of a unit's, such as a country or a higher education institution's, publications that are among the most frequently cited 10, 5 or 1 per cent of a database. You can also examine research collaborations by looking at how often various organisations or countries co-publish with one another.

Subject classification. A common starting point for research assessment is comparing like with like. For peer review of applications, the applications are divided by research area so that a physics-related application is compared with other physics-related applications. These are then assessed by experts in the field of physics. In bibliometric databases, research is divided into a number of subject areas. There is no universal system for subject classification of publications, meaning that the systems vary between different international databases. In the Research Council's database, Thomson Reuters classifies each issue of a journal as belonging to one and six of approximately 250 available subject areas. Articles are tagged with the same subject area(s) as the journal issue in which they are published. Articles in journals which are classified by Thomson Reuters as multidisciplinary are reclassified by the Research Council based on the contents of their reference lists and based on which articles cite them.[5] A drawback which subject classification shares with all disciplinary divisions of research is that it is hard to identify subjects or fields which are not defined in

[5] See *Subject classification of publications in the ISI database based on references and citations* at www.vr.se for an in-depth description of the reclassification process.

the database. For example, interdisciplinary themes can be hard to study. One way of identifying these types of areas is to utilise keywords and in that way identify the publications that should be included.

Fractionalisation. The Research Council generally bases its bibliometric indicators on articles that are fractionalised by subject and address. An article which is classified as belonging to several subject areas in a database can either be counted as a whole article in each subject (*whole counts*), or be divided in equal parts into each of the subject areas (*subject fractionalisation*). In the same way, articles with several authors or author addresses can be counted in *whole counts*, i.e. each author or address is given full authorship of the article. Or the article can be divided into *author fractions* or *address fractions*.[6] There are several reports which argue in favour of fractionalisation (for example [2], [6], [9]) while it is hard to find anyone arguing against it. The most important reason for the Research Council's use of fractionalised articles is that the sum of all fractions, across all subject areas and addresses, corresponds to the actual number of published articles. Using *whole counts*, the same article can be counted several times, and the sum of all subjects and authors (or addresses) therefore overestimates both the total scientific production and the impact. A simple example is shown below.



A country has five universities (A – E), each of which has produced 100 scientific articles. Each such article has been cited exactly once. All five universities have also contributed to 100 collaborative articles. These have each been cited five times each. In other words the total production is 600 articles, together generating 1,000 citations. This gives an average of 1.67 citations per article.

Using address fractionalisation, each university may claim a fifth of the collaborative articles and related citations. For an individual university this means 120 articles with 200 related citations, giving an average of 1.67 citations per article. The same as the actual average.

Using whole counts, each individual university may claim 200 articles and 600 related citations. When totalled across all universities, each collaborative article (and their related citations) will then be counted five times. This would correspond to a total research production of 1,000 articles and 3,000 citations. The average citation count, using whole counts, as well as the average for each individual university, is therefore three citations per article. Significantly above the actual average.

If we wish to compare a university's production and impact with actual average values, for example for a country as a whole, we therefore need to use fractionalisation. Otherwise everyone is above average.

[6] Most publications in the Research Council's database from before 2008 lack links between authors and addresses. The Research Council therefore bases its address fractionalisation on the number of addresses linked to a publication, without regard for how many authors are listed for each address. This can affect the results for individual persons and small research groups, but for sufficiently large publication sample sizes this type of skewing of results will be marginal. The Research Council's address fractionalisation also disregards the order of authors as there are different traditions regarding the order in which authors are placed in an article.

Field normalisation, self-citation, negative attention and sleeping beauties. The Research Council uses mean normalised citations.[7] This means that the number of citations for each publication is compared to a global *field reference value*, which is simply the average number of citations for a publication in the same subject area, in the same year and of the same article type.[8,9] In this way you can identify frequently cited articles in all areas, regardless of citation traditions. You can also calculate the mean normalised citation for a country's total research production and compare it to that of other countries. The Research Council does not include self-citations when calculating the mean normalised citation rate. It is obviously not wrong to cite your own publications, but self-citations do not help give an accurate estimate of a publication's impact, i.e. how much attention it garners from the rest of the scientific community.[10] Also, the fact that a publication is cited by other researchers does not necessarily mean that the attention received is positive. Studies have shown that approximately 10 per cent of all citations are negative in character, with some variation between research areas [4]. However, when comparing larger publication volumes it is reasonable to assume that the units studied have a roughly equal share of negative citations. Another indicator of (or lack of) scientific impact can be gained by studying what share of a total research production goes completely uncited, excluding self-citation.[11] Most citations of an article are made in the first years after its publication. However it is also possible to imagine that some articles are only "discovered" after a few years, and that they subsequently become highly cited. These articles are usually referred to as the *sleeping beauties* of research. In the Science Council's database there are 7 million articles published between 1985 and 1994 which, five years after being published, had less than 5 citations. Of these, roughly 2,000 articles, corresponding to 0.03 per cent, have been cited more than 100 times as of 2013. In other words, we can conclude that sleeping beauties are very rare.

Publication channels' impact. The Research Council uses various bibliometric indicators to assess the impact of a research production. One indication of impact, besides citations, is that the results are accepted by distinguished and widely circulated scientific publication channels (journals, book publishers etc.). The publication channel's impact does not necessarily tell us anything about the impact of individual articles, but combined with a larger publication volume it serves as a good indicator. The indicator can also be used for all research areas where there is reliable information about publication channels. An indicator based on publication channels should not be conservative and should be able to adapt to changes in publication patterns, such as the increased publication in open access journals.

[7] The mean normalised citation corresponds to the Mean Normalised Citation Score (MNCS) which is used at the Centre for Science and Technology Studies (CWTS) at the University of Leiden [8].

[8] The frequency of citations differs between types of articles, for example, a review article is generally cited more than a normal article. The Research Council's publication database therefore differentiates between normal research articles and review articles.

[9] Field normalisation can occasionally be problematic if a small studied publication sample (for example for an individual person or research group) differs markedly from the field's general publication traditions.

[10] In the Research Council's database, a citation is considered a self-citation if one of the author names (last name and initials) in the citing publication is found among the author names of the cited publication.

[11] The share of uncited publications is steadily declining. 30 per cent of articles published in the early eighties were uncited, while the corresponding number for articles published between 2000 and 2005 is 20 per cent. This is mainly due to reference lists becoming increasingly longer. They have increased from an average of 16 references in the early eighties to an average of 26 references for articles published between 2000 and 2005.

The Research Council currently uses mean citation rate to quantify the impact of a journal.[12] In our neighbouring counties of Norway, Denmark and Finland, a model is used where a scientific committee divides all scientific publication channels into two levels, each giving different publication types different scores.[13] Such a system is not dependent on citation information but requires a lot of work to assign levels to the various publication channels. Compared to a journal's average citation, which can be between zero and a hundred citations, a two-level division is quite heavy-handed.

A sufficiently large publication size. What constitutes a sufficiently large publication sample size for bibliometric analyses varies based on what questions are to be answered and how the results are to be used. From experience, the recommendation for aggregate citation-based indicators tends to be a research production where the sum of all fractionalised articles corresponds to at least more than 50 articles. This amount can be lowered to more than 20 articles if the results shall be used by subject experts, along with other information or if the analysis is limited to a specific research area. An important limiting factor when it comes to publication sample size is the generally uneven distribution of citations between articles. There are only a small number of publications that receive a high number of citations, while the majority receive few or no citations. This means that a research production's mean citation rate is dominated by a few well-cited articles, and that most publications are cited less than the average. This means that if the publication sample size is too small the concept of "mean citation rate" becomes unstable. Small sample sizes can also contain varying proportions of negative citations and are therefore harder to compare. In certain cases a small specialised publication sample can deviate significantly from the field's general publication traditions, which affects the mean normalised citation value (see footnote 9). Combined, this means that the Research Council only uses bibliometric comparisons between individuals as a supplement to other information during peer review conducted by subject experts.

Bibliometrics on an individual level and the h-index. Citation data on an individual level should only be used by subject experts and mainly as a supplement to a list of publications. A common bibliometric indicator on an individual level is a researcher's h-index[14], which measures a researcher's productivity and the impact of their research production. It is an indicator which is easy to produce and relatively easy to understand, but which can also be blunt and misleading. An h-index of 10 can for example mean that a researcher has 10 articles, each of which is cited 10 times, or that a researcher has 10 articles which are each cited 100 times. The indicator is not field normalised, which means that it is easier to achieve a high h-index in areas with a high frequency of publication and where each other's work is often cited. This means that comparisons, even among closely related fields, can become problematic. Additionally, the h-index often includes self-citations and is also strongly tied to a person's career age, as researchers over time accrue an increasing amount of publications which can be cited. To correct some of these issues, different versions of the h-index have been created. For example, Publish or Perish[15] uses ten different versions of the

[12] The most common indicator for describing a journal's impact is the *Impact Factor* that is produced by Thomson Reuters, based on the journals' previous citations in their database *Web of Science*. While a journal's mean citation rate is based on the actual number of citations it has received, the *Impact Factor* uses a journal's previous citations to forecast how it will be cited in the future. A journal's impact is not always a stable indicator. It has been shown that journals with less than 50 articles per year changed their *Impact Factor* by an average of 50 per cent between 2002 and 2003 [1].

[13] See http://www.regjeringen.no/upload/KD/Vedlegg/AU/Statsbudsjettet2014_UH.pdf for more information about the Norwegian indicator for scientific publication.

[14] h-index or *Hirsch index* was introduced in 2005 by Jorge E. Hirsch as a way of characterising individual researchers' scientific production [3].

[15] www.harzing.com/pophelp/metrics.htm.

h-index to describe a researcher's impact. The assessment of the Research Council is that the mean normalised citation and the proportion of top 10 publications are better and more transparent indicators of impact, and it therefore does not use the h-index or similar indicators.

Bibliometrics and the evaluation of applications. The Research Council's evaluation panels annually assess and prioritise around 6,000 applications by scientific quality, originality and the applicants' expertise.[16] The bibliometric data that is included when making assessments varies between different calls for proposals and different research areas. In cases where bibliometric data is included in the grounds for the assessment, this is only used by subject experts as part of a larger collection of factors. Bibliometric data is never the only factor used to assess an applicant's merits.

Since the autumn of 2014, the Research Council has an application system, PRISMA, which allows for standardised bibliometric data to be used in the evaluation of applications, as the applicant's list of publications is stored in the database and relevant publications are appended to the application by the applicant. The Research Council is therefore working to, based on the aforementioned guidelines, produce joint bibliometric data for the evaluation process. The data shall be produced in the same way for all research areas with a similar degree of coverage in the Research Council's bibliometrics database. The proposed bibliometric data will be available in PRISMA at the earliest in 2016.

If and how the data shall be used as part of the assessment work will be specified in the instructions that apply to each respective call for proposals/funding form and research area. When using bibliometric data in the evaluation of applications, it shall always be understood that the evaluation panels' subject experts are to only use the data as part of a larger set of data and with great care. *Researchers' merits shall never be assessed, compared or ranked solely based on bibliometric data.* Furthermore, bibliometrics used in the evaluation of an application shall only be based on the publications in PRISMA which the researcher has presented in the application's list of publications. This means that any time constraints or instructions to highlight a limited number of publications in the application also limit the bibliometric data. The bibliometric information should be seen as a complement to the list of publications, to help the evaluators assess the publications' scientific impact. In the Research Council's bibliometric data, each publication that is linked to an application will, if present in the Research Council's bibliometric database, be supplemented with information on the number of authors, number of full citations and the mean normalised citation. It is also shown if the publication is among the top ten per cent of highly cited articles in its subject area. For publications which are not included in the Research Council's bibliometric database, no citation information is provided. The evaluation panels receive no information about, and should not use, the h-index.

Continued bibliometrics development. The importance of bibliometrics as a monitoring tool has increased which has also led to a development of the area as a field of research. As previously mentioned, bibliometrics is most useful when analysing large publication volumes. For this reason, the development in the field has been influenced by, among other things, network analysis and Big Data research. One example is the new tools that are used to cluster and visualise underlying structures in large bibliometric networks.[17] Such tools can be used to show how different research areas cite each other and how new research areas are developed over time. The Research Council hopes that bibliometrics continues to develop as a method of analysis, and that its areas of application become both broader and deeper.

[16] For details regarding assessment of applicants' expertise, see the evaluation handbook for each respective research area at www.vr.se.

[17] See for example mapequation.org.

References

[1] Adler, R., Ewing, J., Taylor, P. (2008) *Citation Statistics - A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)*, p. 8.

[2] Aksnes, D.W., Schneider, J.W., Gunnarsson, M. (2012) *Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalized counting methods*, Journal of Informetrics, Vol. 6, Issue 1, p. 36–43.

[3] Hirsch, J.E. (2005) *An index to quantify an individual's scientific research output*, Proceedings of the National Academy of Science, Vol. 102, Issue 46, p. 16569-16572.

[4] Moed, H.F. (2005) *Citation Analysis in Research Evaluation*, Springer, p. 84.

[5] Olesen Larsen, P.V. (2010) *The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index*, Scientometrics, Vol. 84, Issue 3, p. 575–603.

[6] Persson, O., Glänzel, W., Danell, R. (2004) *Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies*, Scientometrics, Vol. 60, Issue 3, p. 421–432.

[7] Swedish Research Council (2009) *Bibliometrisk indikator som underlag för medelsfördelning*, www.vr.se.

[8] Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S., Van Raan, A.F.J. (2010) *Towards a new crown indicator: Some theoretical considerations*, arXiv:1003.2167 [cs.DL].

[9] Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., Van Leeuwen, T.N., Van Raan, A.F.J., Visser, M.S., Wouters, P. (2012) *The Leiden ranking 2011/2012: Data collection, indicators, and interpretation*, Journal of the American Society for Information Science and Technology, Vol. 63, Issue 12, p. 2419–2432.