Bibliometric handbook

for Karolinska Institutet

Catharina Rehn and Ulf Kronman, Karolinska Institutet University Library

Version 1.05

2008-12-15

Table of Contents

What is bibliometrics?

One of the earliest definitions of bibliometrics describes it as "the application of statistical and mathematical methods to books and other media of communication" (Pritchard, 1969).

Today, bibliometrics is often used to assess scientific research through quantitative studies on research publications. Bibliometric assessments are based on the assumption that most scientific discoveries and research results eventually are published in international scientific journals where they can be read and cited by other researchers.

Evaluative bibliometrics – "quantitative measurements of qualitative aspects (such as 'quality' or 'reputation') of the science system" (Theodorus Nicolaas Van Leeuwen, 2004) is based on the assumption that the number of citations to a journal article can be considered to reflect the article's impact on the scientific community.

The term *bibliometric indicators* is often used for the results of a bibliometric analysis. One of the definitions of the term *indicator* in the Oxford English Dictionary is "That which serves to indicate or give a suggestion *of* something; an indication *of*" (*Oxford English dictionary* 2000). This draws attention to the fact that the results describe a reality that is too complex to be measured merely by statistics or numbers. In a glossary produced by the Joint Commission on Accreditation of Health Care Organizations, USA, there is another definition that seems close to how the word indicator is used in bibliometrics ("Sentinel Event Glossary of Terms", 2006): "1. A measure used to determine, over time, performance of functions, processes, and outcomes. 2. A statistical value that provides an indication of the condition or direction over time of performance of a defined process or achievement of a defined outcome."

Analyzing bibliographic references

This section contains several examples of the kind of results you can get from a statistical analysis of standard bibliographic references like the one below.

Annu Rev Med. 2006;57:119-37.
Pharmacogenomics and individualized drug therapy.

Eichelbaum M, Ingelman-Sundberg M, Evans WE.

Dr Margarete Fischer Bosch Inst Clin Pharmacol, Stuttgart, D-70376 Germany
Karolinska Inst, Div Mol Toxicol, IMM, Stockholm, SE-17177 Sweden
St Jude Childrens Hosp, Memphis, TN 38105 USA

Pharmacogenetics deals with inherited differences in the response to drugs. The best-recognized examples are genetic polymorphisms of drug-metabolizing enzymes, which affect about 30% of all drugs. Loss of function of thiopurine S-methyltransferase (TPMT) results in severe and life-threatening hematopoietic toxicity if patients receive standard doses of mercaptopurine and azathioprine. Gene duplication of cytochrome P4502D6 (CYP2D6), which metabolizes many antidepressants, has been identified as a mechanism of poor response in the treatment of depression. There is also a growing list of genetic polymorphisms in drug targets that have been shown to influence drug response. A major limitation that has heretofore moderated the use of pharmacogenetic testing in the clinical setting is the lack of prospective clinical trials demonstrating that such testing can improve the benefit/risk ratio of drug therapy.

MeSH Terms:
Biotransformation/genetics
Cytochrome P-450 Enzyme System/genetics

Glucuronosyltransferase/genetics
Humans
Methyltransferases/genetics
Polymorphism, Genetic/genetics
Receptors, Adrenergic, beta-2/genetics
Sodium Channels/genetics

Publication year

An analysis of publication years can for example show trends in how much a unit publishes compared to the rest of the world, or to similar units.

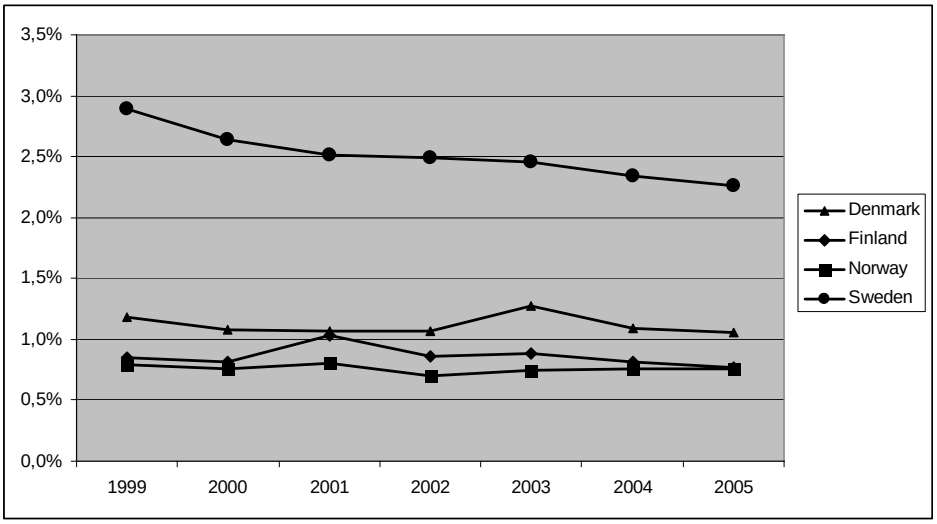

Figure 1. The relative number of articles from Sweden, Denmark and Finland in immunology journals in relation to the world averages of articles in the same field (1999-2005).

Journal title

An analysis of journal titles can for instance give an overview of the publication pattern of a certain unit.

Journal	**Publications**
BONE MARROW TRANSPLANTATION	216
SCANDINAVIAN JOURNAL OF IMMUNOLOGY	173
JOURNAL OF NEUROIMMUNOLOGY	154
JOURNAL OF IMMUNOLOGY	148
CLINICAL AND EXPERIMENTAL IMMUNOLOGY	136

Table 1. The five most frequent immunology journals used for publication 1995-2005 by authors at Karolinska Institutet.

Author names

It may be of interest to identify prolific authors in a specific country or at a specific unit.

Author	**Publications**
Link, H	123
Ringden, O	114
Ljungman, P	93
Bjorksten, B	89
Holmdahl, R	86

Table 2. The five most prolific Swedish journal article authors within the field of immunology 1995-2005.

Additional information can also be had by analysing the co-publication patterns of authors in a specific area.

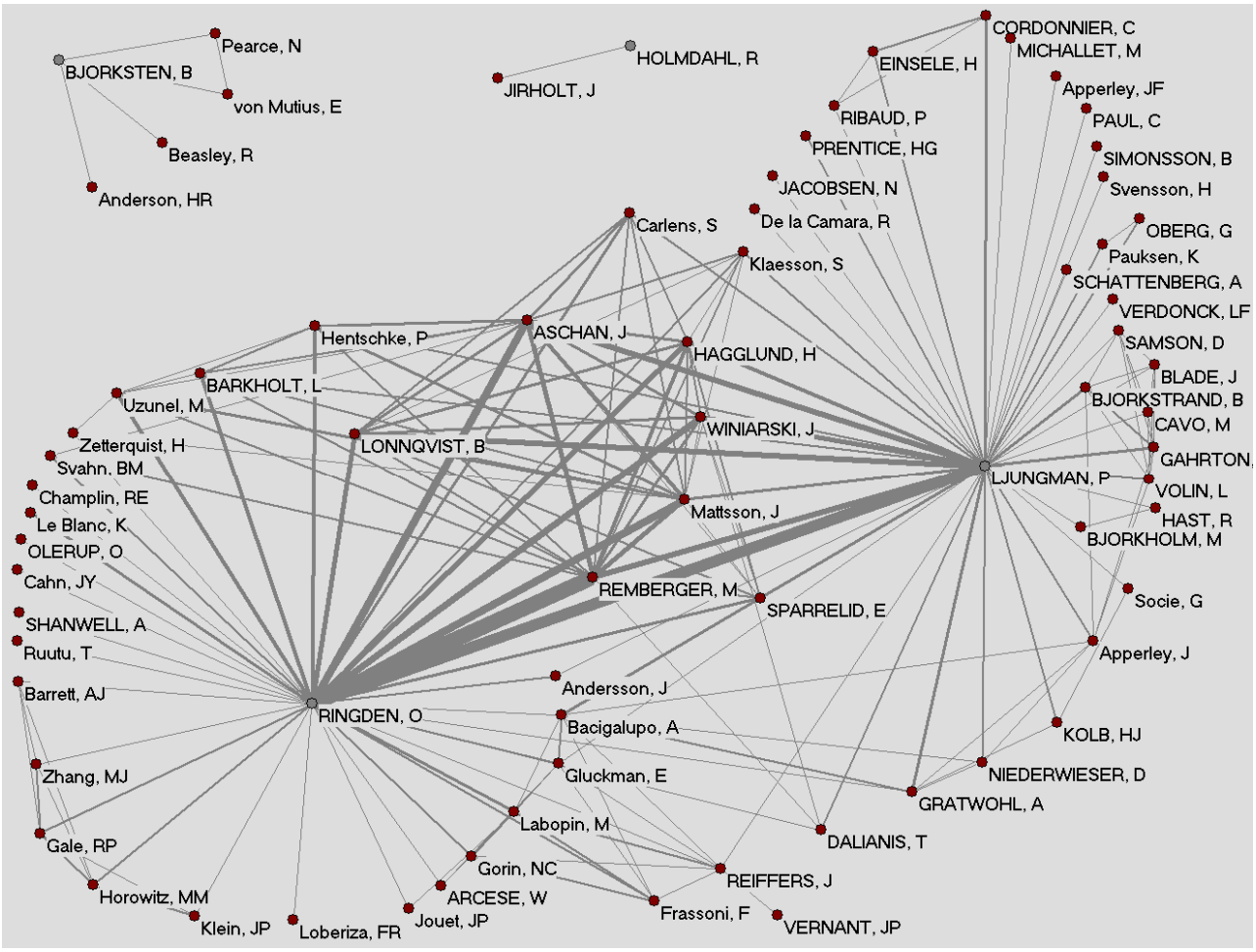

Figure 2. The co-publication pattern of the five most prolific Swedish journal article authors that publish in immunology journals (1995-2005). Each author included in the analysis has a connection to at least 15 other authors. The connections shown are at least 5 publications strong (suggesting a rather strong connection between the authors) and an author shown has at least one connection of that strength. One of the five authors listed in table 2 is missing from the picture. H Link did not have at least 15 connections to other authors. This may either suggest that he does not co-publish much or that H Link is in fact several authors with the same name.

Author addresses

Analogous to an analysis of author names, an analysis of author addresses can identify prolific countries, universities or other organizations and give an overview of the co-publication patterns.

Country	**Publications**
Sweden	447
Denmark	209
Finland	151
Norway	150
Iceland	20

Table 3. The number of immunology publications in 2005 from the five Nordic countries.

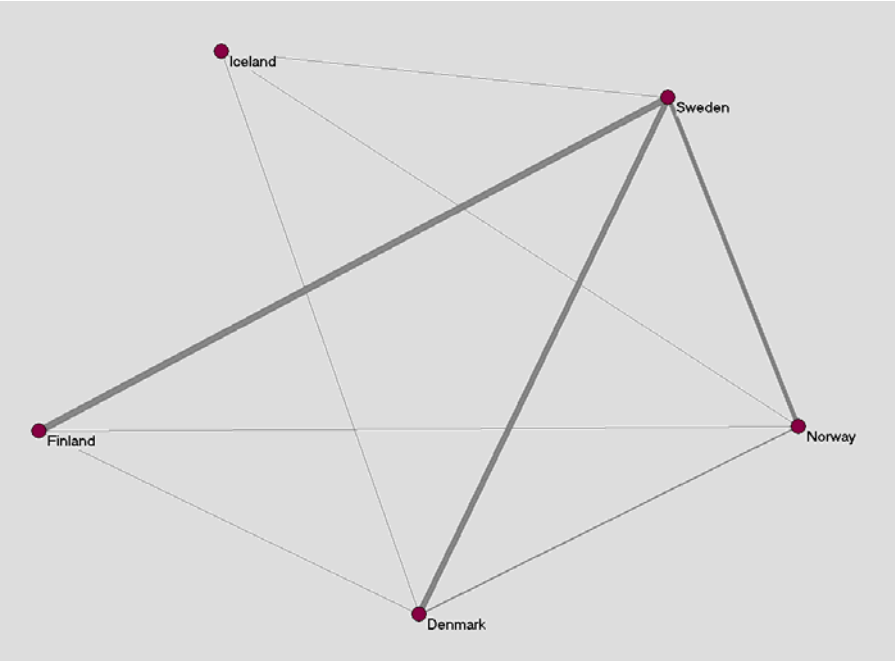

Figure 3. Co-publication patterns in immunology between the Nordic countries in 2005 based on publications in immunology journals.

Keywords

If you have adequate keywords assigned to the publications it is possible to study which subjects appear often in the publications of a unit and connections between them, a so called co-word analysis. This kind of analysis is easier to make if the reference contains terms from a controlled vocabulary like MeSH (Medical Subject Headings from the National Library of Medicine, USA).

Keyword	**Publications**
T-Lymphocytes	204
Antibodies, Monoclonal	92
Spleen	77
Lymphocyte Activation	76
Epitopes	68
B-Lymphocytes	67
Lymphocytes	66
Cell Line	62
HIV-1	61
Cytotoxicity, Immunologic	52

Table 4. The ten most frequent keywords in publications by one specific author.

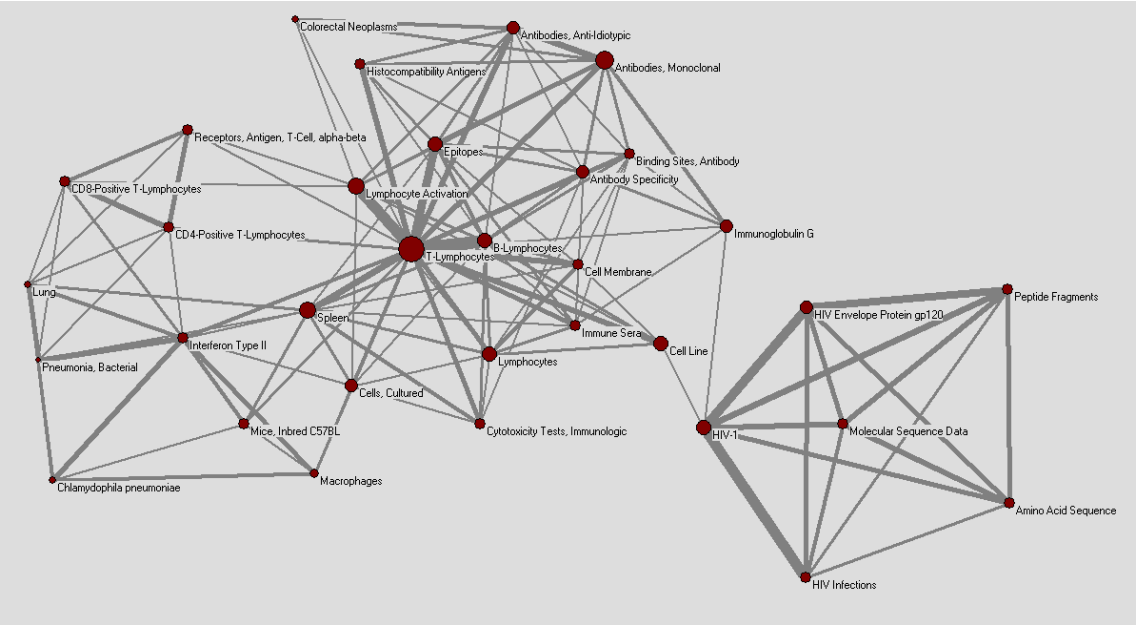

Figure 4. The co-occurrence pattern of MeSH-terms in publications by one specific author (MeSH check tags excluded). To be included, the MeSH-term has to be connected to at least five other MeSH-terms and only connections stronger than 10 co-occurrences show up as lines. It is possible to see a strong core of immunological terms surrounded by three other foci (HIV1, bacterial lung infections and colorectal neoplasms).

Combinations of the above

The analyses above are often improved by combining more than one aspect.

By combining author names or author addresses and keywords it is possible to localize prolific authors or organizations in a specific field. This also makes it possible to identify authors that are connected through common subjects – something that may identify existing research networks or the possibility of a new research network.

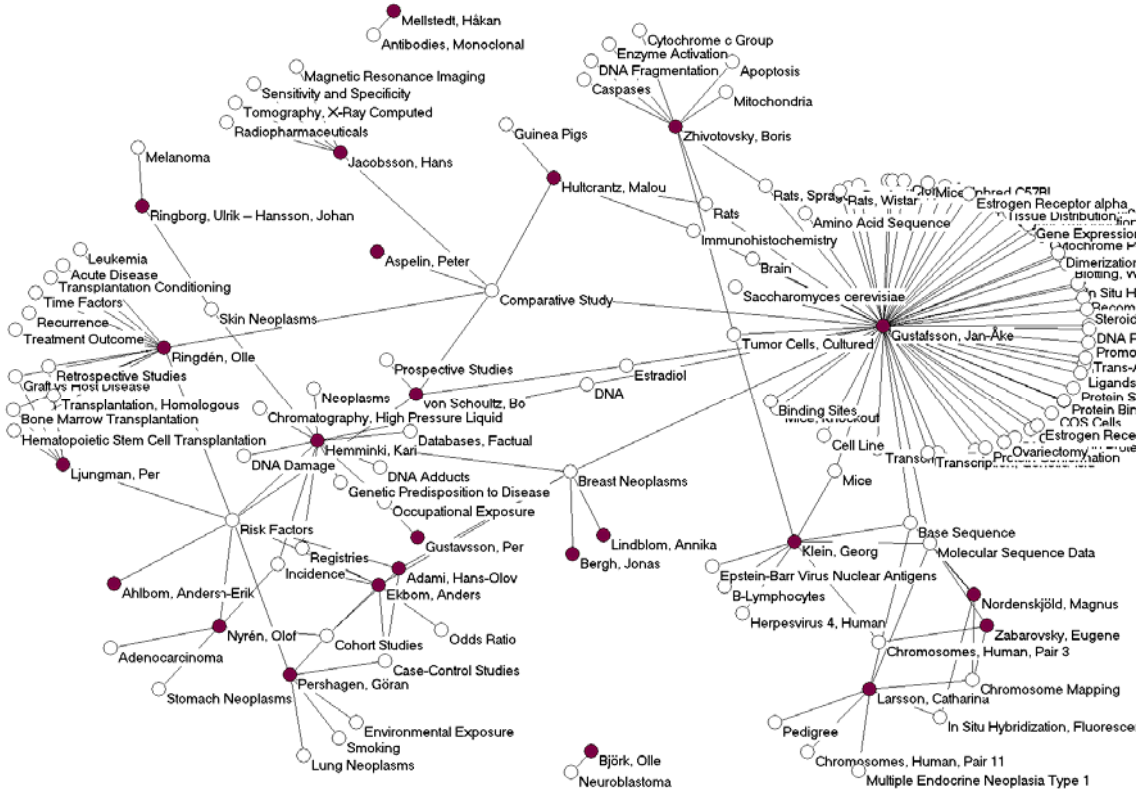

Figure 5. This network shows connections through subjects (represented by MeSH-terms) between research groups at Karolinska Institutet that have published in the field of oncology.

Analyzing reference lists and citations

When bibliographic records include reference lists to cited articles, we can extend our statistical analysis on the connections between various publications. In general, the additional information obtained from reference lists can supply at least two more aspects to a bibliometric study:

- The possibility to find publications in the same area by identifying the publications that cite (refer to) or are cited by the publications that you already have identified.
- The possibility of a bibliometric quality assessment.

Bibliographic coupling

A bibliographic coupling analysis connects publications that share items in their reference lists (see figure 6), that is, refer to the same publications.

The assumption behind bibliographic coupling is that publications within the same subject share core material and the more like the publications, the more like the reference lists. One specific trait of this method is that it makes it possible to find conceptual connections between articles that are so new that they haven't had the time to be cited yet.

An example on how bibliographic coupling can be used is the case where you find a highly relevant article that refers to an important previous article. You may then do a bibliometric search to se what other articles that refer to this previous article to se if there are newer relevant articles on the same subject as the first article.

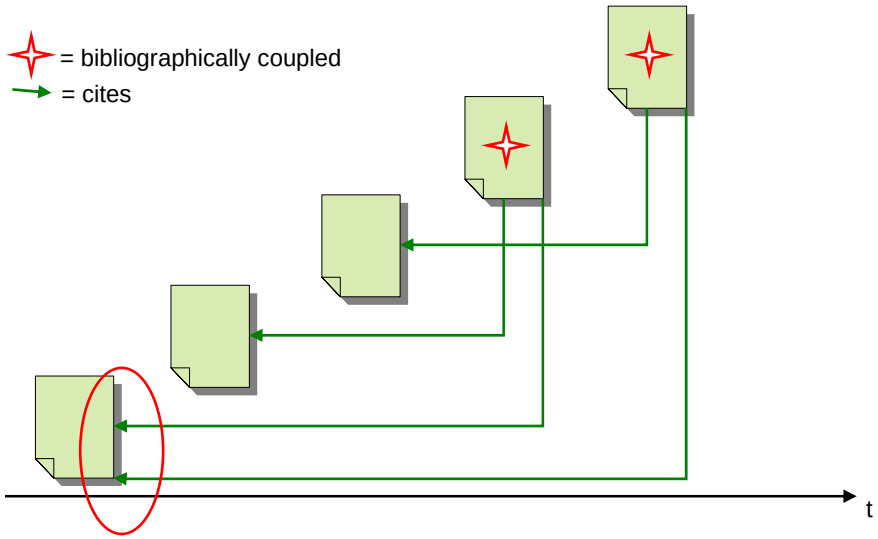

Figure 6. Bibliographic coupling analyzed via shared references. The more recent star-marked articles may be subject-related to each other, since they refer to the same older article.

Co-citation analysis

A co-citation analysis studies reference-pairs, i.e. papers cited (referred to) by the same publication (see figure 7). By doing a co-citation analysis you may find older articles that are related to each other, even thought they don't refer to each other.

This type of analysis will usually generate clusters with highly cited articles since two highly cited papers are more likely to be co-cited in several reference lists than two lowly cited ones.

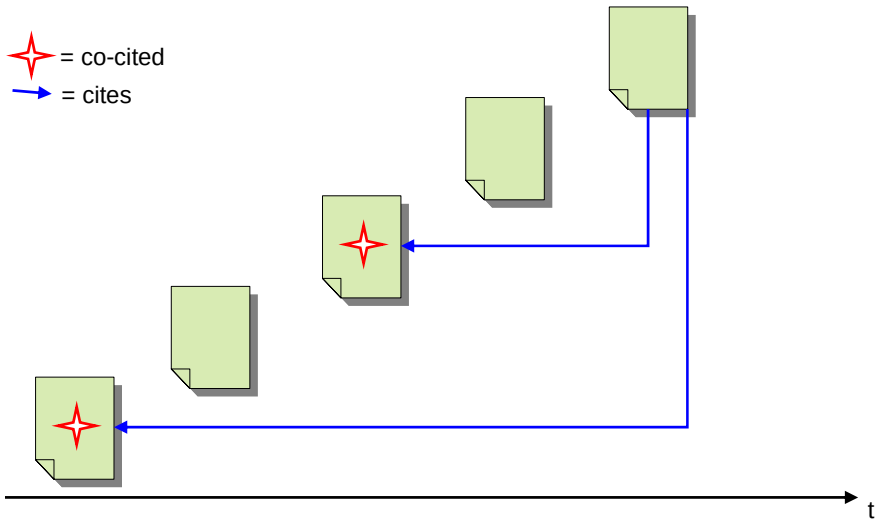

Figure 7. Co-citation analyzed via shared reference lists. The star-marked articles may have something in common, since a more recent article refers to them both.

Quality assessment

If citation rates of a unit's publications are high compared to the expected citation rate per publication this shows that the unit's articles have had significant impact on the scientific society, which in turn can be assumed to indicate that the articles are of high quality.

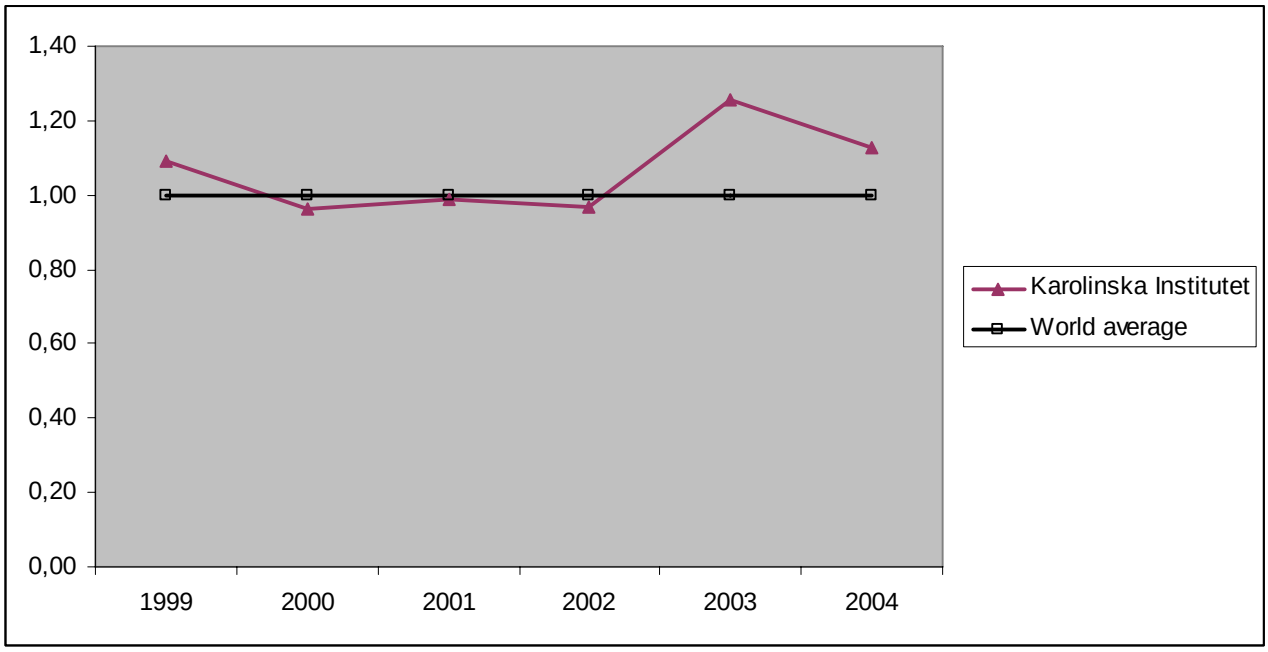

Figure 8. The mean value of the number of citations to Karolinska Institutet articles in immunology journals in relation to the world average (=1) of citations to articles within the same research area. Karolinska Institutet articles are cited above average from 2003 and onwards, which may be an indication of improved article quality during recent years. The final interpretation, however, has to be made by experts in the field of immunology.

How do you get cited?

Before getting further with the discussions of citation-based science performance indicators it may be in place to take a look at the nature of scientific citations.

A comprehensive summary of research on why authors cite each other has been written by Henk Moed (Henk F. Moed, 2005). It has been found that it is reasonable to assume that most citations are ”positive”, that is to say a sign of the fact that the citing author finds something useful in the material he cites.

Deviating citation patterns, such as negative citations, can affect an analysis of an individual article or author, but this adverse effect tends to disappear in an analysis of larger aggregations of authors, such as departments, universities or countries.

The number of citations to a publication is affected by (Glänzel, 2003, p. 61):

1. The subject matter, and within the subject, the “level of abstraction”.
   The publication activity in theoretical fields (e.g., mathematics) and in engineering is lower than in experimental fields or in the life sciences. Articles in a research field with the custom to write long reference lists also receive on average more citations.

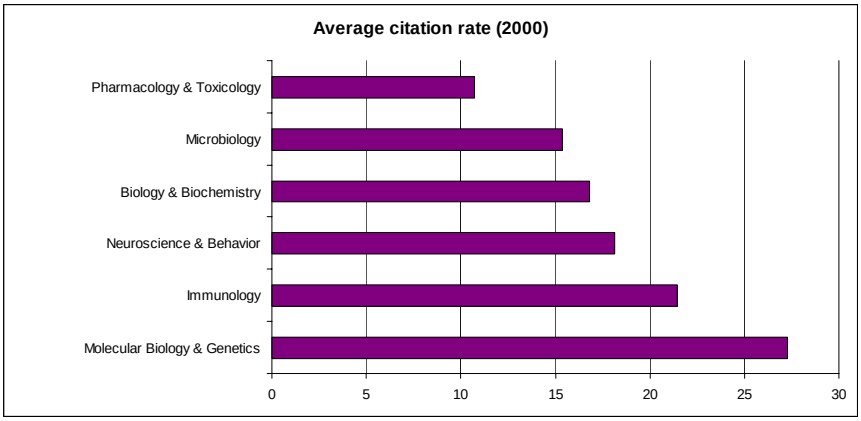

Figure 9. The average citation rate in 2000 for six different fields of research.

2. The age of the publication.
   Older publications have a longer time period during which they can receive citations. Older articles are therefore on average more cited than new ones.

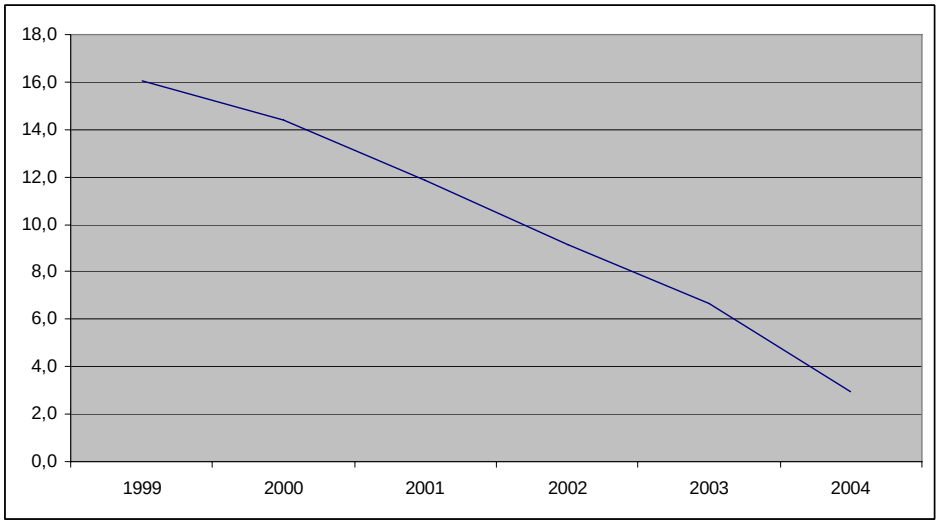

Figure 10. Average cumulative citation rates recorded in the Thomson citation indices 2005 for publications published in 1999-2004 in the field of immunology.

3. The "social status" of the publication (through the author(s) and the journal).
   At higher aggregation levels (e.g., at institutional or national level), the influence of the factors regarding author age and social status tends to vanish since populations at this level are rather heterogeneous.

4. The document type.
   Certain types of papers, such as review papers, tend to be cited more than original articles. In most bibliometric studies only some document types are included. "Only those document types that are conveyers of relevant scientific information are taken into consideration. Such publications are, in particular, journal papers of the type research articles, letters, notes and reviews. Meeting abstracts, editorial material, corrections/errata, retractions, book reviews and other document types not listed above are only objects of special bibliometric studies." (Glänzel, 2003, p. 46)

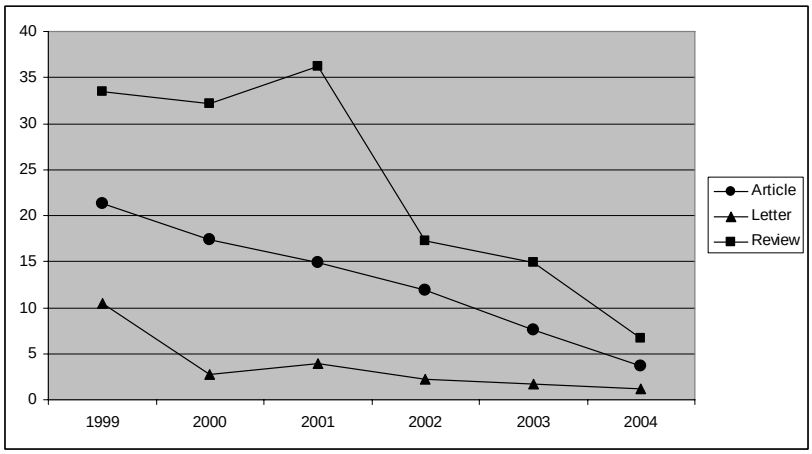

Figure 11. Average cumulative citation rates for articles, letters and reviews in the Thomson citation indices 2005 for items published 1999-2004.

5. The length of the observation period.
   Different research areas reach their ”citation peaks” in different years after publishing. It is therefore important to decide not only the publication years of the papers you wish to study (the ”time window”) but also for what years after publication you wish to count the citations (the so called ”citation window”).

In a few cases, the citation pattern differs from the one described above:

- The Mendel effect or the “sleeping beauty”.
  A single paper remains uncited for a long time until the rest of the research community discover its value and start citing it.
- "Obliteration by incorporation”.
  The information in a publication has been used so much that it’s considered public knowledge and as such uncited. This effect takes very long to appear, and most bibliometric analysis only includes the last 5 or 10 years.

However, both these varieties are very rare and do not affect an analysis made on a sufficiently large number of publications.

The distribution of citations to a group of articles is nearly always skewed, even if you take the above mentioned factors into consideration. Some of the publications in a group are much more cited than others. The distribution within the same group can vary from uncited articles to articles among the 1% most highly cited compared to the world average. This can be seen both in leading groups and in less prominent groups and it is also true for papers in high as well as for those in low impact journals. Moed says that this may be because some papers can be considered as ‘flags’ and others as ‘bricks’. The bricks lay the foundation that the flags need to stand on in order to be high quality papers, but only the flag papers get cited. (Henk F. Moed, 2005, p. 86).

What do bibliometric analyses measure?

Bibliometric analyses result in *indicators* of research quantity and performance. They can also provide measurements of connections between researchers and research areas through statistical analysis of co-publications and citations.

Below is a list with examples of various indicators. For more indicators and full descriptions and definitions of indicators; see the appendix: *Bibliometric indicators – definitions and usage at Karolinska Institutet.*

Quantity indicators: Number of publications and citations

Examples:

- **Number of publications and citations**. The two most basic bibliometric indicators describe the number of publications and citations attributed to a group of authors (a research group, a department, a university or a country) during a specified time period.
- **Number of publications and citations per researcher** is a relative measure. It compensates for the size of the studied unit and therefore indicates scientific output in relation to invested resources.
- **World share of publications.** The unit's number of publications in relation to the world production.
- **Number of publications in Thomson ISI indices.**
- **Number of publications in top-ranked journals.**

Performance indicators: Normalized citation counts

Examples:

- **Field normalized citation score (including the "crown indicator")** measures the research impact of a group of authors. It compares the average number of citations to the group's publications to the average number of citations to international publications from the same year, in the same subject area and of the same document type.
- **Top 5 %** shows the share of publications attributed to a group of authors that belong to the 5% most cited publications in the world from the same year, in the same subject area and of the same document type.
- **Journal normalized citation score.** Expresses how much a unit's publications are cited in relation to other articles in the same journals that they are published in.

Journal performance indicators: Impact indicators

- **The ISI journal impact factor** for a scientific journal is a mean value that corresponds to how many times an average article published in the journal has been cited.
- **Journal to field impact score** is an alternative to the ISI impact factor. It measures the average level of citations to articles in one journal in relation to other journals within the same scientific field and can be calculated for different time spans.
- **Normalized journal impact** is normalized on an individual article level and corresponds to the field normalized citation score calculated for publications in one specific journal. Each publication in the journal in question is compared to other publications from the same year, of the same document type and published in journals in the same fields.

Structural Indicators: Publication and citation patterns

Structural indicators are for example the fields in which a unit publishes and the fields in which it is cited. Further one can make descriptions of the cognitive structure of the unit's research field, or of co-authors and the co-author's affiliations (organizations, countries etc).

One example is the use of connection maps to illustrate how much different units publish together or how a selected number of units are connected through a common field of research.

How do you perform a bibliometric analysis?

Applied bibliometrics, as it is used today, analyzes the number of scientific articles published by a selected number of authors, citations to these articles and connections between articles, authors and subjects.

Getting data

The most common way to get data for a bibliometric analysis is to extract the information from an already existing database containing bibliographic information. The source is often one or more of the citation indices made available by Thomson Scientific, but it can also be a locally produced database of the publications from one specific unit or indeed any database containing information about the publications that are to be included in the analysis.

Thomson Scientific ISI citation indices

Most bibliometric analyses use data originating from one or more of the three ISI citation indices supplied by Thomson Scientific. (ISI – the Institute for Scientific Information – founded by Eugene Garfield in 1958 is now a part of Thomson Scientific.)

The most important citation index for medicine, life science and the natural sciences is the Science Citation Index Expanded (SCIE). This contains references to articles from more than 5 900 scientific journals ("Thomson Scientific: Citation Products"). There is also a Social Sciences Citation Index and an Arts and Humanities Citation Index.

Some of the advantages of the Thomson citation indices are:

- Multidisciplinary
- Go many years back
- Contain all author addresses
- Contain citation data
- Include full journal content – not just parts
- Reasonably standardized

Including all three indices, Thomson Scientific indexes about 8 500 of an estimated number of more than 22 000 active, refereed scientific journals (Ulrichsweb.com).

Since Thomson Scientific use the reference lists from publication records in their own indices to select what journals to include, it is reasonable to assume that the Thomson citation indices contain the most cited and most important academic journals.

Subscribers to the Thomson citation indices can, for example, access them though the web based service Web of Science. This is a relatively easy-to-use search interface that provides the opportunity to create lists of publications and citations attributed to researchers, research groups, departments, universities or countries. It’s however not suited for more complex bibliometric analyses, including the calculation of mean values or connection mapping. For this you have to purchase or download data from the Thomson citation indices and use locally developed applications for the calculations.

As a consequence of the strong bibliometric focus on data from the Thomson citation indices, most bibliometric indicators are reliable only in research areas where publishing in scientific journals is the main mode of communication. This is often the case in natural sciences, technology and medicine, but analyses of areas within the humanities or social sciences must apply other methods as well (Henk F. Moed, 2005, p. 42).

PubMed/Medline

Medline is the world's largest medical database. It is produced by the National Library of Medicine (NLM), USA and covers the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. Medline contains bibliographic citations and author abstracts from more than 4,800 biomedical journals. The database at present contains over 14 million citations dating back to the mid-1960. Coverage is worldwide, but most records are from English-language sources or have an English abstract ("PubMed Overview ", 2006). A lot of work is put into the indexing of article references with the controlled vocabulary MeSH.
PubMed is NLM:s own search interface to Medline. In addition to Medline this retrieves:

- "OLDMEDLINE for pre-1966 citations.
- Citations to articles that are out-of-scope (e.g., covering plate tectonics or astrophysics) from certain MEDLINE journals, primarily general science and general chemistry journals, for which the life sciences articles are indexed for MEDLINE.
- In-process citations which provide a record for an article before it is indexed with MeSH and added to MEDLINE or converted to out-of-scope status.
- Citations that precede the date that a journal was selected for MEDLINE indexing (when supplied electronically by the publisher).
- Some life science journals that submit full text to PubMedCentral® and may not have been recommended for inclusion in MEDLINE although they have undergone a review by NLM, and some physics journals that were part of a prototype PubMed in the early to mid-1990's."

("What's the Difference Between MEDLINE and PubMed? Fact Sheet", 2006)

Medline does not include any information about citing or cited references.

Chemical Abstracts

Chemical Abstracts contains references (including some information about cited/citing references) to publications in biochemistry, organic chemistry, macromolecular chemistry, applied chemistry and physical, inorganic, analytical chemistry. It contains references to articles and other document types (from 1907-) from more than 9000 major scientific journals, including cover-to-cover coverage for nearly 1,500 key chemical journals (since October 1994). It also includes patent (and patent family) references from more than 50 active patent-issuing authorities around the world ("SciFinder - Content at a Glance - Patent and Journal references ", 2006).

Information about cited references is included since 1997 and this citation information is available on records from 1990 and onwards (Roth, 2005).

Scopus

Scopus is a relatively new database produced by Elsevier Science Publishers. It covers over 15,000 peer-reviewed journal titles from more than 4,000 international publishers and includes 245 million references from the reference lists of the publications covered by the database. Titles from all geographical regions including non-English titles are considered for inclusion as long as English abstracts can be provided.

Scopus covers several different disciplines such as Chemistry, Physics, Mathematics and Engineering, Life and Health Sciences (including references retrieved from Medline), Social Sciences, Psychology and Economics, Biological, Agricultural and Environmental Sciences and General Sciences. ("Scopus Info - Scopus in Detail - What does it cover?" 2006)

In Scopus, reference lists are included from 1996, and these are automatically matched to corresponding publication records to obtain the citing/cited information of included publications.

We expect that bibliometricians may want to try and use data from Scopus for their analyses, but no studies using Scopus data have been published yet.

Google Scholar

Google Scholar which is produced by Google Inc. also contains a kind of citation information. However, the information about citing references is rather unreliable and may contain both duplicates and mismatches.

Local repository

A local repository usually does not contain any citation information, but it can contain other information that can be used in a bibliometric study. Depending on the policy on entering publication information, a local repository may well be the most complete record of publications from that particular research unit.

Selecting a unit of analysis

The starting point in a bibliometric analysis is to select a group of publications, usually on the basis of information available in the Thomson citation indices. This selection of publications forms the unit of analysis.

The publications may for example be selected on the basis of the authors' organizational affiliations and can theoretically be:

- Author
- Research group
- Department
- Research Centre/Network
- University
- Country

A substantial amount of local data preparation and verification is necessary in order to create a unit of analysis based on a research group, a department, a research centre or a research network. This information is very difficult to locate in data from the Thomson citation indices and may in many cases not be present at all. It is even difficult to attribute publications to a particular university since both organization names and their addresses may be written in many different ways and two different universities occasionally share a common name.

A unit of analysis can also be selected based on the properties of individual articles (instead of authors or author affiliations).

- Individual publications
- Journal
- Subject – often based on subject classification of the journal
- Document type – article, review, note, letter, conference proceeding, etc.
- Publication year

Since statistical methods are used in bibliometric research, the results improve with larger units of analysis. This is partly because isolated phenomena – as negative citations – are cancelled out by the large amount of articles (Henk F. Moed, 2005, p. 80). Using bibliometric indicators based on

any unit of analysis that contains less than 10 articles (as an individual researcher or article) is not to be recommended ([Anonymous], 2005, p. 10).

It is also necessary to take into consideration any possibility of a systematic bias. This could for example be different citing traditions or conventions for including and ranking authors that vary significantly between different research areas. (Henk F. Moed, 2005, p. 223).

Choosing time and citation windows

When you do a bibliometric study, you have to decide what time intervals you will use as the basis for your data collection process.

First, you have to decide the *time window*, i.e. the time over which you want to study the unit's publication performance, i.e. the years in which the studied articles were published.

Second, you have to decide between which years the subsequent articles that are going to deliver citations to the analyzed articles have to be published to be included in the citation count. This is called the *citation window*.

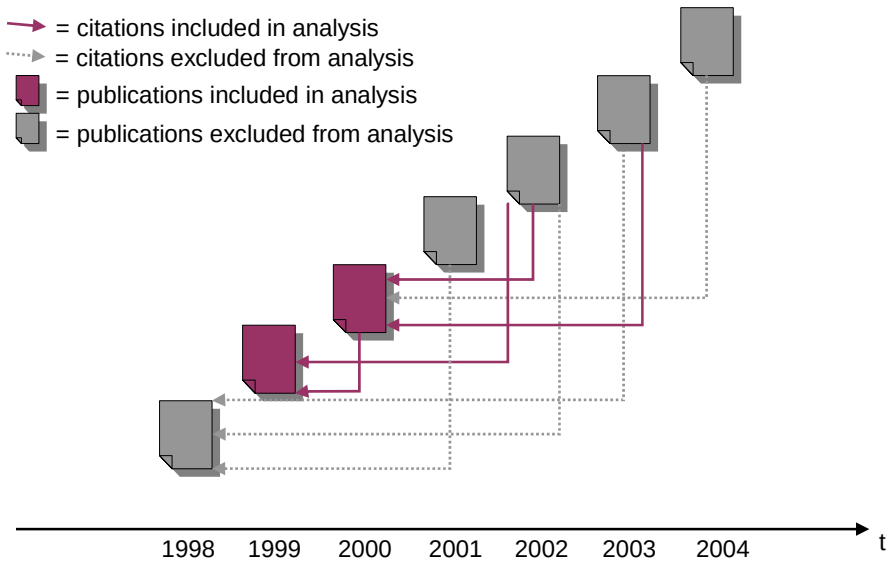

Figure 12. The time window (publication years for cited publications) is 1999-2000 and the citation window (publication years for referring (citing) publications) is 2000-2003.

Time window

A time window of 8-10 years (two PhD student generations) can be considered sufficient to analyze the publication activities of one individual research group (H. F. Moed, 2002, p. 38). When analyzing larger units, the importance of the length of the time window diminishes, but it should of course be the same for all analyzed units.

Citation window

There are a few different types of citation windows, and they will affect the analyses results differently.

A citation window can either be fixed or variable. A fixed citation window is set to a specific time period, for instance 3 years to include the citation peak of most disciplines. A variable citation window uses differently sized citation windows for publications from different years, the citations to all publications are for example counted up to and including 2004, regardless of publication year.

Fixed citation windows are usually “overlapping”, that is, they depend on the year of publication. If an analysis uses fixed, overlapping citation windows of for example three years, citations to a publication from 1999 would be counted 1999-2002 and citations to a publication from 2000, 2000-2003. This gives a somewhat fairer image if you wish to compare citation counts of publications from different years, but it would be better to use normalized citation counts (see below).

A fixed non-overlapping citation window would have the same “citation years” for all publication years included in the specified time window, for instance, citations from 2000-2003 will be counted both for publications from 1999 and 2000. This is of course a disadvantage for the more recent articles when it comes to citation counts.

A long fixed citation window will exclude newer publications since these will not have had time to reach the end of their citation window. A short fixed citation window may cause several normalization groups to have too few citations to be suitable for analysis.

When using an indicator that is normalized with regard to publication year, the citation window will by default be variable. If you plan to use basic indicators however, you have to decide if you wish the citation windows to be fixed and overlapping/non overlapping or variable.

Technical issues with data

Misspellings

Any type of misspelling in the data, be it of author names, addresses, journal titles or other information, will lead to incorrect numbers of citations and publications. This type of error is fairly common and may be both due to misspellings by the original authors or by mistakes made by the database producer. In large amounts of data however, the effect of these errors is often negligible.

Unknown addresses

As much as 2.4% of the articles, letters, notes and reviews in the Science citation index 1993-2003 and 14% in the Social Science Citation Index (excl. SCI) lack information about authors and author addresses. (Henk F. Moed, 2005, p. 186) The percentage is higher if other document types are included. These references will automatically be excluded from many types of analyses.

Connections between references and the corresponding article

Automatic matching of reference lists to corresponding articles always fails to identify some of the connections. Moed says: ”when data are derived from ’simple’ or ’standard’ citation-matching procedures, citation statistics at the level of individuals, research groups, journal and countries are strongly affected by sloppy referencing, editorial characteristics of scientific journals, referencing conventions in scholarly subfields, language problems, author-identification problems, unfamiliarity with foreign author names and ISI data-capturing conventions.” (H. F. Moed, 2002). Moed found the overall number of discrepant cited references in the Web of Science to be about 7% but it may be much higher in specific situations.

Aspects of interpretation

Several aspects of an essentially non technical nature have to be considered before it is possible to perform a bibliometric analysis, or indeed, interpret the results of one.

Identifying an organization as the unit of analysis

It is not always easy to define an organization through the author addresses. The addresses are written by the authors themselves, and it is not uncommon for them to write either the department name or the university name instead of both. In some cases only the main address of a great consortium of writers is written on the paper, and the addresses of the individual authors are left out. Many organizations also have several different units with separate addresses, such as a university, the attached medical school and the university hospital. (van Raan, 2000) De Bruin and Moed have described some of their work on unifying addresses in (Debruin & Moed, 1990).

In fact, manual identification seems to be the only way to get reliable data on the publishing organizations. And even then, you need to decide if the publications of a university are to be defined as the publications of the people attached to the organization or the publications that were published while someone was working within the organization.

This decision depends on whether you wish to assess what the unit has already achieved (a retrospective study) or what it has the potential to achieve (a prospective study).

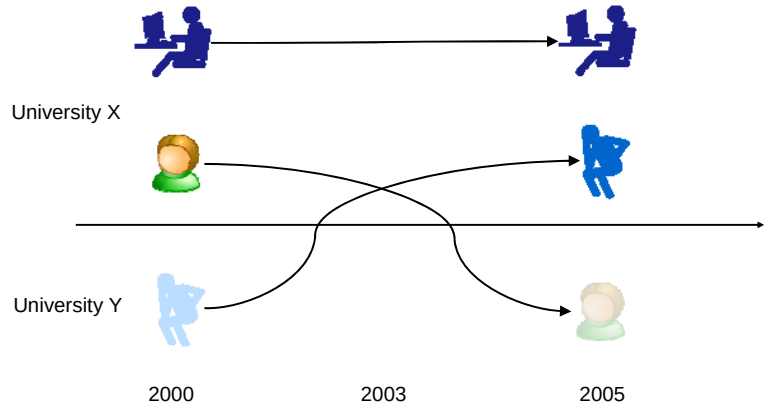

Figure 13. A retrospective study only includes articles published during the time an author was working at the analyzed unit. Articles published before moving to a unit or after moving from a unit are disregarded.

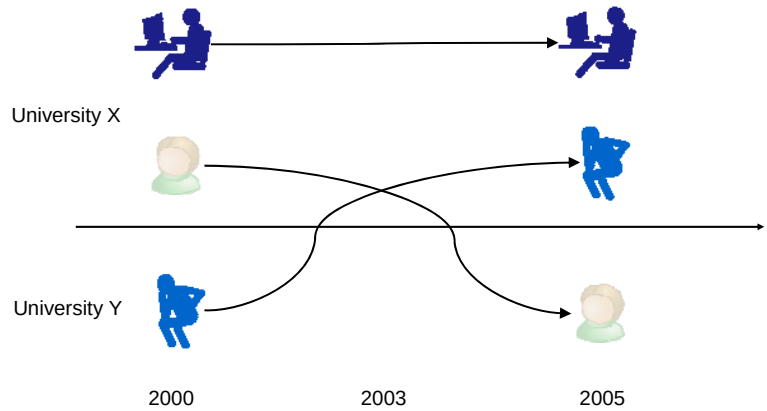

Figure 14. A prospective study includes all articles published by authors still working at the analyzed unit, including articles published before moving to the unit. Articles published by authors that have moved from the unit are disregarded, even if they were produced while working at the unit.

The retrospective study will depend on correct identification of the organizations in the address field, the prospective study on the identifications of the names in the author field.

Categorizing publications into research fields

Most bibliometric analyses and indicators use some kind of categorization of the analyzed publications into research fields. This is usually based on the subject headings of the included articles or journals supplied by the database producer. At present, most bibliometric analyses do not have any information about the subject of an individual article, but only the subjects of the journals that the articles are published in, and this is used as a proxy for "article-subject". Often, the ISI classification scheme for journals is used, since this is readily available and covers many different disciplines. It would however be desirable to use a standardized classification scheme on the individual publication level, and the Karolinska Institutet bibliometrics group is currently examining the possibility of using the Medical Subject Headings ("Medical Subject Headings", 2006) which are available for all publications covered by Medline.

Self citations

The expression "self citations" can be used both for an author or a unit citing their own papers and for publications in a journal citing publications in the same journal. The first of these is the more common usage, and the one used in this handbook (unless specified otherwise).

Studies have shown that self citations do not significantly influence analysis results when you study a sufficiently large number of publications (Glänzel, 2003, p. 57). This is probably because most researchers refer to their own work in equal quantity as a natural part of scientific communication. On group level however, small differences in citation counts may indeed influence indicators. The small number of publications increases the possibility of extreme indicator values.

It is speculated that authors in the United States cite themselves more than the average, but this hypothesis has yet to be proven.

You can address the aspect of self citation by:

- trying to exclude self citations when calculating indicator values
- noting them so that the interpretation of the indicators can be affected by the amount of self citations
- assuming that they are evenly distributed and hence ignoring their effect when calculating the indicators

It is very difficult to remove self citations when calculating indicators and it requires data from a comprehensive citation database such as the Thomson citation indices and usually also verification by the analyzed unit.

The aspect of self citations has been left out in the indicator descriptions in the appendix *Bibliometric indicators – definitions and usage at Karolinska Institutet*.

Fractionalization and Weighting

Not all authors can be considered to have played an equally large part in producing a publication and some publication types may be considered to be of more importance than others in a particular research area. There are therefore different ways of compensating for this and similar factors when calculating the number of publications or number of citations attributed to a certain author or unit.

Fractionalization means that only a part of a puclication and/or a citation is attributed to the unit of analysis according to some mathematical frequency principle.

Weighting means that some types of publications and/or citations are considered to be of more importance than others and therefore given a higher "weight". Implementation of weighting introduces a subjective element into the analysis since the factor by which the publication/citation

is to be weighted has to be decided by the analyst in collaboration with experts in the research fields that are to be evaluated.

The impact of fractionalization and weighting on indicators may be considerable, and one should always take into account what kind of behaviour will benefit or be at a disadvantage with the different ways of counting and take this into consideration when studying the indicators.

Fractionalization and weighting according to the number of institutions in the affiliation list may for example discourage inter-institutional cooperation, something that may not at all be a desired effect.

Fractionalizing and weighting the number of publications

If you do not fractionalize or weight the number of publications in any way, each publication is attributed to each of the authors and each publication will be considered to be of equal value. An article with five authors will for example be attributed once to each of the five authors and thus counted as "published" five times, which means that the sum of all articles by all individual authors will be larger than the total number of individual articles. This way of counting is often called *full* or *integer* counting.

Fractionalizing by authors or producing units

If you wish to compensate according to the composition of the author list there are several alternatives:

- Fractionalization by giving each of the authors an equal part of one publication, if for instance an article has five authors, each author can count 0.2 as his/her article. The same principle can be used when attributing publications to institutions instead of authors.

Weighting by authors or producing units

- Weighting according to place in the author list. In some disciplines, for example medicine, the author that has done most of the work is often first in the author list and the most senior researcher is last. You could for example give the first author 50% of the publication, the last author 20%, and divide the remaining 30% between the rest of the authors. However, this differs not only between disciplines but also between countries, departments and journals and this makes comparison with other units rather uncertain.
- Weighting according to number of occurrences in affiliation when attributing a publication to an institution. It is then assumed that the institution with several authors has contributed more to the work in producing the publication.

Weighting by document types

In some areas, especially those where other types of publications than journal articles are important, you can also consider giving more weight to some publication types, for instance: Ptotal_weighted = number of books * 5.0 + number of refereed journal articles * 1.0 + number of book chapters * 0.8.

Fractionalizing by field

In you are doing an analysis that produces separate indicators for each included scientific field, the number of publications can be fractionalized between different fields. A publication categorized as belonging to both oncology and haematology will for example contribute half a publication to each field.

Weighting by field

If a publication can be considered to belong more to one field than to the others that it has been classified with, this field can receive a higher share of the publication than the other fields. This is very difficult to implement with a categorization made with the ISI journal subject categories but

may be possible when using for example MeSH terms for article classification since these include a division of terms into major, ordinary and subheadings.

Fractionalizing and weighting the citation count

Fractionalizing and weighting by authors (or other producing units), document types and fields

The citations to an article can be fractionalized and weighted with the same methods and considerations as the number of publications.

Fractionalization according to length of reference list

When an article refers to another article, it can be considered that the value of the referral stands in inverted proportion to the length of the reference list. In other words, a referral from an article with a long reference list could be considered to be of less value than a referral from an article with a short reference list.

This aspect can be imposed by fractionalizing the value of citations by dividing the citation value with the number of items in the reference list. For instance, a citation from an article with 20 references would give the referred item a citation value of 1/20 = 0.05.

Weighting according to citation source

Citations to an article may come from many different sources, and the sources cannot always be considered to be of equal importance. It may be desirable that citations from articles in journals with high impact indicators should be weighted higher than those from articles in journals with low impact indicators.

A special kind of citation weighting according to source is the Google PageRank ranking algorithm. In short, this algorithm gives higher weight to citations from publications that have themselves a high citation count. There have been some successful experiments that try this algorithm for weighting of the importance of citations to journal articles, (Chen, Xie, Maslov, & Redner, 2006) something that must be considered an interesting development.

Language

An analysis of Web of Science data shows that non-English publications on average get cited much less than publications in the English language. The indicator values of non-English authors improve when you exclude their non-English publications (van Leeuwen et al 2001). Excluding non-English publications may thus be an option when using bibliometrics to assess the research performance of non-English researchers.

Citing traditions

The citing traditions within different fields affect the number of citations given to different articles. Different fields also vary in how quickly a paper will be cited, how long the citation rate will take to peak and how long the paper will continue being cited. Eugene Garfield describes several factors that contribute to a fields “citation potential”. One factor is the length of the reference lists, which for example is twice as long in biochemistry as in mathematics. Another factor is the size of the fields “core literature” (Garfield, 1979, p. 248). Many disciplines connected to medicine and life sciences have a high citation potential.

It is possible to get an overview of the citing traditions in a particular field by looking at the reference lists for a selection of articles within that field. The length of the reference lists, the subjects of the cited publications and their publication years to some extent give an insight in the citation speed and behaviour of that particular field.

Skewed distribution of citations

The distribution of citations to publications is by no means linear, even for the articles of one single author. We have already mentioned Henk Moed's theory about "brick" and "flag" papers and a closer description of this theory is available in his book "Citation analysis in research evaluation" (Henk F. Moed, 2005, pp. 216-218). For many statistical calculations however, you need a curve that is at least approximately linear. This can partly be achieved by using a logarithmic scale for one, or both, of the diagram axes. For more information, see (Seglen, 1992).

Coverage

The quality of a bibliometric analysis improves with increasing coverage of publications in the area you wish to study. Moed has studied the coverage of the Thomson Citation Indices and got the following results (Henk F. Moed, 2005, p. 42).

- Excellent:
    - Molecular biology & biochemistry
    - Biological sciences related to humans
    - clinical medicine
    - physics & astronomy
    - chemistry
- Good:
    - Applied physics & chemistry
    - Biological sciences primarily related to animals and plants (including agriculture)
    - Engineering
    - Geosciences
    - Mathematics
    - Psychology
    - Psychiatry
    - Other social sciences related to medicine and health
    - Economics
- Moderate:
    - Sociology
    - Education
    - Political Sciences
    - Anthropology
    - Arts & Humanities

Is a bibliometric analysis advisable for a particular unit?

Before a bibliometric assessment is undertaken, there are some things to consider:

- You have to estimate how large a share of the unit's publications that can be found in the databases available to you, and if this share is sufficiently large to make bibliometric analysis a reliable option. Is for example the international English serial literature the unit's

main mode of communicating its findings? If not, the analysis cannot be made using an already existing citation database, and indicators based on citations can therefore not be included.

- Is it possible to, manually or automatically, identify the papers that belong to this particular unit in the database you chose as your data source? The smaller the number of publications for a unit, the larger the consequences if one or a few of these publications are missed.
- Is the total number of publications by this unit sufficiently large to produce reliable indicators? A small number of publications can produce very extreme indicator values.
- Who will receive the results of the analysis? Interpretation of the indicators can be difficult and continuous discussions with the person or persons requesting the analysis are necessary in order to agree on the best indicators in the individual case and what these may show, and indeed, not show, with regard to the analyzed units.

Henk Moed supplies information on coverage indicators and how to assess a unit with insufficient coverage in the ISI citation indices in his book ”Citation Analysis in Research Evaluation” (Henk F. Moed, 2005).

Choosing bibliometric indicators

Many bibliometric researchers stress the importance of not considering the results from any bibliometric analysis to be “truths”. Bibliometric methods contain so many simplifications that they only supply a very limited picture of the research they are trying to describe.

No bibliometric indicator should be put to isolated use. Several indicators should always be combined to achieve a more comprehensive picture of the scientific production of a unit (T. N. van Leeuwen, Visser, Moed, Nederhof, & van Raan, 2003). A field normalized citation mean indicator should for example always be accompanied by a so called top indicator that shows if the mean value of citations to the unit’s publications is due to a few very highly cited articles or a majority of publications cited a bit above average, and by a quantity indicator to show how many publications that are included in the analysis.

It is important to see bibliometric indicators as one of several tools to be used by competent reviewers with specific knowledge about the research areas included in the analysis. This is for example evident when publications containing very new or unconventional research results are included in an assessment. These will not yet have been cited, which means that any assessment based solely on bibliometric indicators will not discover the possible potential of the research groups in question.

Bibliometric indicators

This chapter contains some examples of the most commonly used and/or important bibliometric indicators. For a more complete list; see the appendix: *Bibliometric indicators – definitions and usage at Karolinska Institutet*.

Basic bibliometric indicators

Basic bibliometric indicators are characterized by being mainly raw publication and citation counts, where there is no or little compensation for the size of the analyzed unit, nor are the characteristics of citation patterns regarded.

Number of publications and citations

Two very basic bibliometric indicators are the number of publications and citations during a specific time period. These two indicators do not compensate for the size of the publishing unit or the document type of the publications. However, they can be useful to someone with knowledge of the research area under study, especially if the indicators are used to compare similar research units or as a complement to other bibliometric indicators.

Since most publications are written by more than one author, it is not always straightforward how you should count the number of publications for each author. Different kinds of weighting schemes for publications and citations have been described earlier in this handbook.

Number of publications and citations per researcher

Publication and citation counts in relation to the number of active researchers or employees at the studied unit are two somewhat more refined indicators of scientific production and impact. It can however be surprisingly difficult to find out the number of active researchers at one particular unit.

Citations per publication

The average number of citations that articles published by an analysed unit during the analyzed time span has received.

It gives an indication of the average scientific impact of a unit's published articles, but it does not take into account that older articles usually are more cited and that citation rates vary between document types and subject areas.

h-index

The h-index is the number of publications (h), attributed to an analyzed unit during the analyzed time span, that have at least h citations.

It is very easy to calculate in the Thomson Web of Science, but it has the same disadvantage as the other basic indicators – it compares documents of different types, published in different years, in totally different subjects, with each other.

The h-index gives positive bias to senior researchers with older articles, since these have had more time to be cited. However, this effect is somewhat dampened by the fact that any new articles from an author with a high h-index have to have comparable citation levels to keep up the high h-index. Another criticism on the h-index is that it disfavours scientists with a short career, since the h-index never can be larger than the number of published articles, no matter how important and well-cited the articles are.

ISI journal impact factor

The ISI impact factor was designed by Eugene Garfield around 1960 as a means to measure the scientific impact of a specific journal. It gives an average value on how many times an article in the journal has been cited. It is defined as the average number of citations given in a specific year to documents published in that journal in the two preceding years, divided by the number of documents published in that journal in those two years.

The impact factor is used by Thomson to localize the most important scientific journals in each research area and is in this respect also used as a library collections management tool.

The impact factor is often written as a number corresponding to the average number of citations to an article in the journal (including the document types *reviews* and *notes*). The journals with the highest impact factors can reach numbers of around 50, and journals like New England Journal of

Medicine, Nature and Science have impact factors of about 30. Most journals have an impact factor below 1.

The fact that the ISI impact factor is based on citations only 1-2 years old can be considered a compromise between the need of getting a quick appraisal of new journals and letting the publications reach their citation maximum (the year when the publication receives most of its citations). Most articles reach their citation maximum 3-5 years after publication, so that would for many research areas be a preferable citation window.

The citation patterns vary so much between different research areas that the ISI impact factor should not be used to compare the scientific impact of journals in different subjects.

Since the ISI impact factor is relatively easy to find and understand it has become very popular and is often used to assess the quality of articles, researchers, departments and universities by studying the journals they publish in. This is unadvisable and not what the impact factor was intended for. The correlation between the impact of a journal and the citations of the journals individual articles is very poor which makes it impossible to predict the number of citations to an article on the basis of the impact factor of the journal it has been published in.

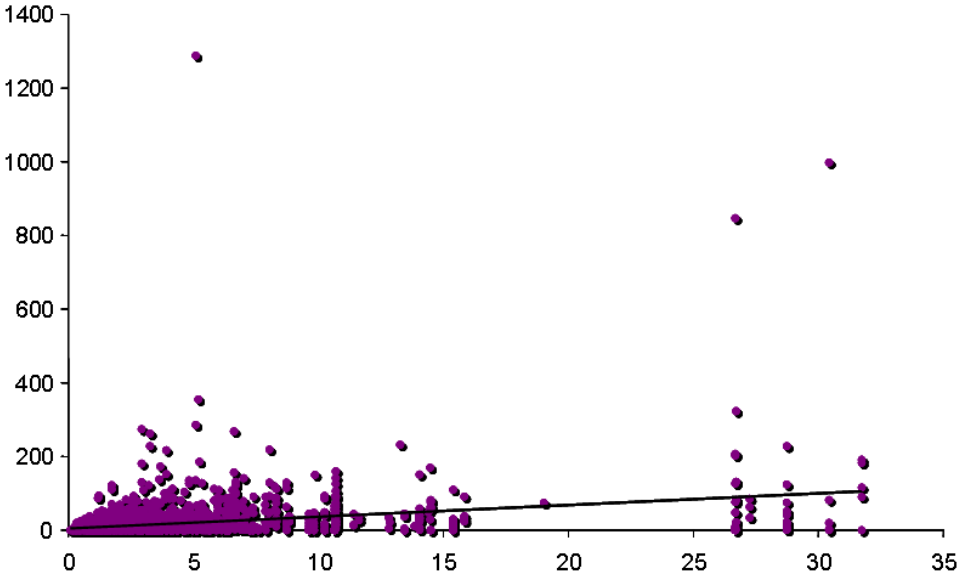

Figure 15. Impact factors of journals compared to individual citations based on articles produced by Karolinska Institutet in 1997. The black line represents the impact factors of the journals and the spots represent individual articles published in the journals.

However, if you wish to study very recently published articles that have not yet been cited, a journal impact indicator may be the only possible performance indicator. The indicator is then based on the assumption that the refereeing process is more rigorous in high-impact journals, which means that only high-quality research will be accepted. For any type of analysis that relates to journal impact, you should however use one of the more advanced normalized journal impact indicators described below, rather than using the ISI impact factor.

Number of publications in high-impact journals

Publications in high-impact journals are often considered to be of high quality. It is not uncommon for researchers to be asked to supply information about their ”mean impact factor”, the average value of the impact factors of the journals they have published in, when they apply for a grant or a new position.

Sometimes a research unit or a university also displays how many publications they have in journals with very high ISI impact factors, the 20 or 40 most highly ranked, as a sign of quality research produced by authors at that unit. This measurement is usually not compensated for the size of the analysed unit, which means that larger institutions get higher values. The research areas of the analysed unit will also affect this figure, so sometimes different journal impact factor lists are used for different research areas.

However, as mentioned earlier, the impact factor of a journal cannot predict the number of citations that any individual publication will receive. Often about 20% of the publications in a journal receive 80% of the citations and many articles are cited 0-1 times even in high impact journals.

Advanced bibliometric indicators

Advanced bibliometric indicators have three important aspects "built-in".

- Publication year – citations accumulate with age which means that older articles are more highly cited.
- Document type – the number of citations to different document types varies significantly. Review articles, for example, generally receive more citations than regular articles.
- Research area – the citation patterns are different in different research areas

The advanced indicators always include a normalization process, i.e. a comparison of publication citation counts to the citation count average of publications of the same document type, published the same year, in the same subject.

Field normalized citation score (including the "crown indicator")

The crown indicator compares the average number of citations to the publications of an analyzed unit to the average number of citations to international publications from the same year, in the same research field and of the same document type.

The normalized citation score is usually written as a decimal number that shows the relation to the normalized world average 1, which means that 0.9 shows that the analyzed publications are cited 10% less than the world average and 1.2 that they are cited 20% more.

The field normalized citation score may be calculated in two different ways. The initial field normalized citation score invented by CWTS, called the "crown indicator", aggregates the unit's publications as a whole, calculates a mean citation value, and then divides that with the average of the field citation scores for the fields, years and document types the aggregated articles belong to.

At Karolinska Institutet, an alternative way to calculate this indicator is being developed. Is is called the "Item oriented field normalized citation score average", and it is calculated by normalizing each individual publication's citation rate against an average citation rate for articles in the same subject area, the same type and of the same age, and finally the average of all the normalized citation values are calculated.

Read more about the variants of this indicator in the appendix: *Bibliometric indicators – definitions and usage at Karolinska Institutet*.

Top 5%

Top 5% shows the share of publications attributed to a group of authors that belong to the 5% most cited publications in the world from the same year, in the same field and of the same document type.

The Top 5% indicator is just as the crown indicator written as a decimal number that shows the relation to the normalized world average 1. A value over 1 shows that the analyzed unit has more of its publications among the top 5% than the world average, a value below 1 that it has less.

Top 5% is often used as a complement to the crown indicator. It shows if a high crown indicator value is achieved through a few very highly cited articles or a larger number of articles cited above average. It may also identify highly cited articles from a group with a low crown indicator value whose top publications would otherwise have been unnoticed.

Knowledge in the specific research area is required to decide which of the two publication patterns is the better sign of high-quality research.

Next generation indicators

Current bibliometric indicators have made it possible to relate the number of citations received by a publication to a world average based on subject area, time since publication and document type. This has been a major step towards an increased validity and applicability of bibliometric indicators. However, it should be noted that these indicators are not final, and new indicators are under development. Examples of indicator development currently performed at Karolinska Institutet are:

- New subject classification: Current indicators use a subject classification based on journal subject categories supplied by Thomson. We have previously shown that this subject classification has several shortcomings (J. Lundberg, Fransson, Brommels, Skar, & Lundkvist, 2005). Therefore Karolinska Institutet is testing the possibility to use MeSH-terms as the basis for subject classification on an article level.
- New statistical method: To calculate the state-of-the-art indicator “crown” the received number of citations is divided by the expected number (the world average for publications of the same type, from the same year, within the same area). This does not take the citation distribution of different areas into account. Therefore Karolinska Institutet is exploring the possibility to use other statistical methods, for example z-score (Jonas Lundberg, 2006).

Publication patterns

Co-publication

If the publication data used for analysis contains all author addresses (as in data from the Thomson citation indices) it can be used to find patterns of co-publication between different authors and units. These can be visualized in co-publication tables and maps that show for example the authors, universities and countries that publish together and to what extent.

In most co-publication maps not all authors and connections have been included, since this would make the elements in the map too many for a satisfactory visualization. It is therefore important to know the inclusion criteria for the different elements.

- Number of publications for an author to be included.
- Number of connections for an author to be included.
- Strength of connection for it to show up on the map.

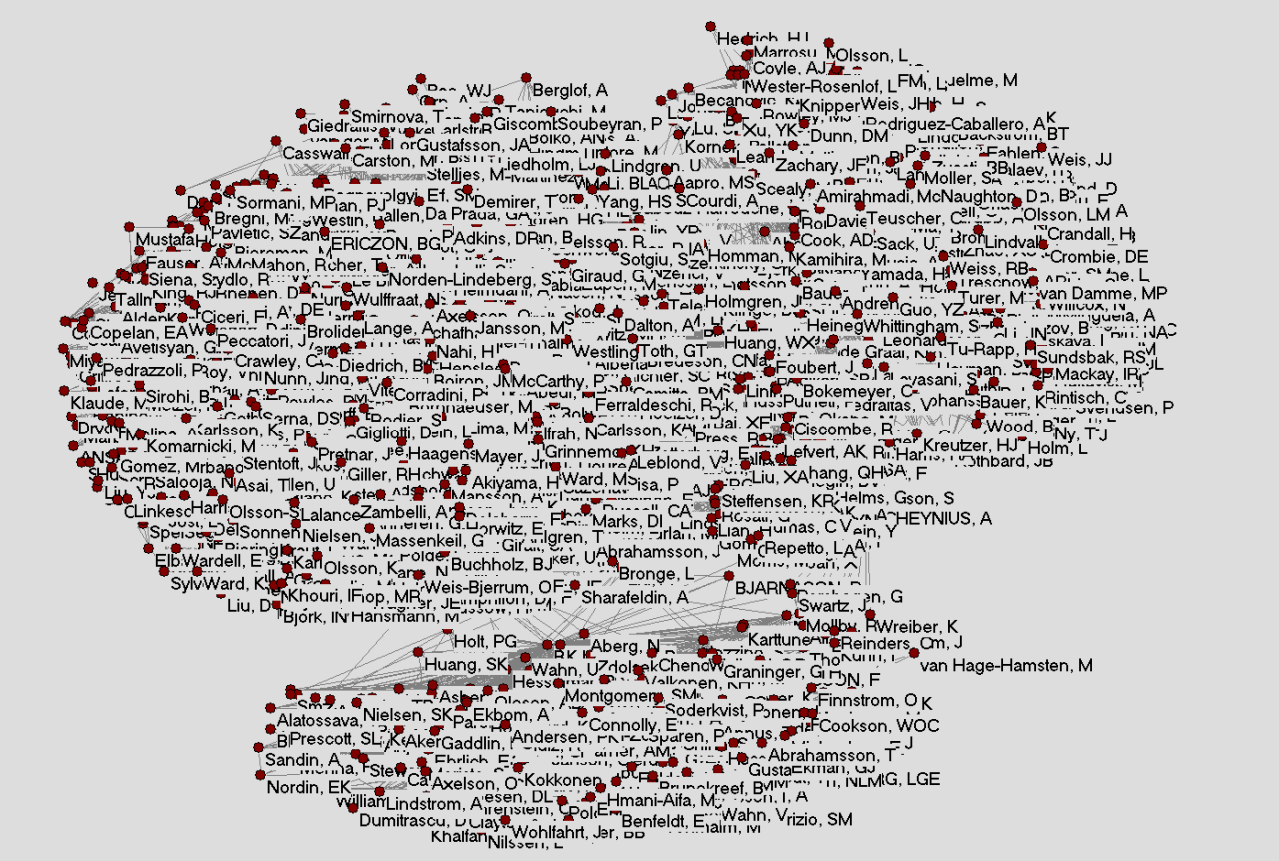

Figure 16. A co-publication map of the five most prolific Swedish journal article authors that have published in immunology journals 1995-2005 (Ringden O, Ljungman P, Bjorksten B, Holmdahl H and Link H). All 1745 authors connected to these five are included in the map which makes it impossible to interpret.

If you choose to have a minimum number of publications for each author you will reduce the number of authors by excluding the non-prolific ones. At the same time you also exclude any connections that the authors still present in the map have to any of the excluded authors.

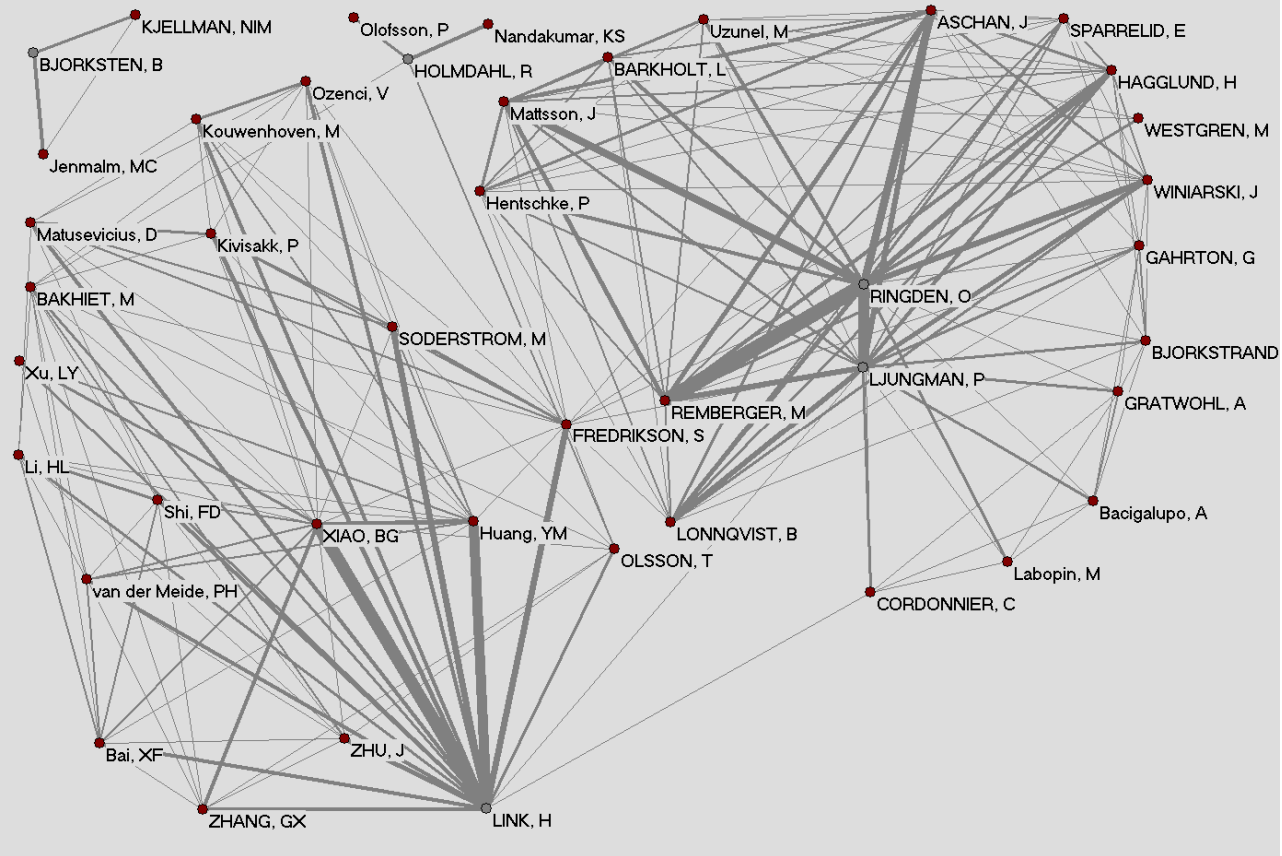

Figure 17. A co-publication map of the five most prolific Swedish journal article authors that have published in immunology journals 1995-2005 (Ringden O, Ljungman P, Bjorksten B, Holmdahl H and Link H). Each author included in the analysis has published a total of at least 15 articles together with one or several of the selected Swedish authors.

If the analysis only shows authors that have many connections to other authors, you will not only exclude non-prolific authors but also authors that only co-publish with one or a few partners.

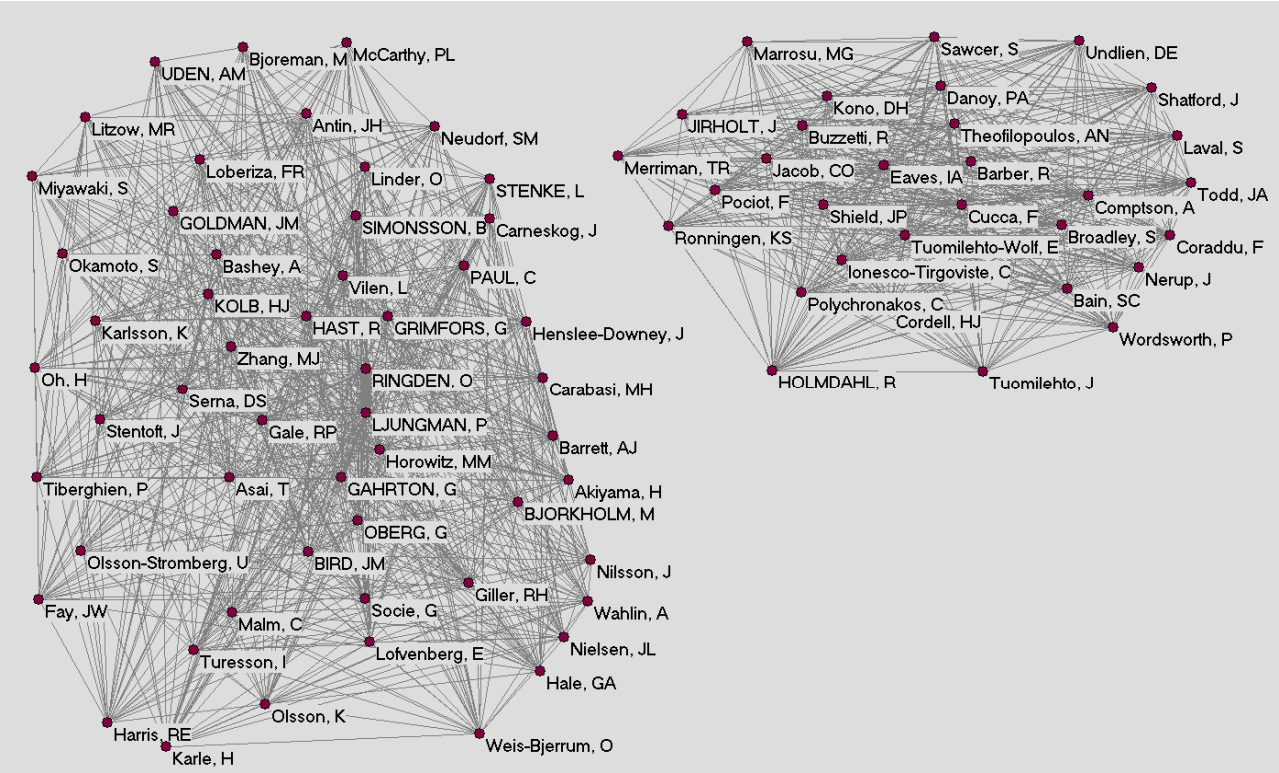

Figure 18. A co-publication map of the five most prolific Swedish journal article authors that have published in immunology journals 1995-2005 (Ringden O, Ljungman P, Bjorksten B, Holmdahl H and Link H). Each author included has published at least 25 articles together with other authors in the network (all 1745). Link and Bjorksten have disappeared from the map due to this inclusion criterion. Ringden and Ljungman have published many papers together and are both in the left cluster. Holmdahl has not published together with Ljungman or Ringden and is in the right cluster. Even though the two clusters seem separate, we cannot draw that conclusion since the original data did not include all immunology publications 1995-2005, only those connected to 5 Swedish authors.

With or without excluding any authors you may also wish to hide connections that are very weak, for example connections of less than 10 publications. That will result in the map showing very strong cooperation patterns.

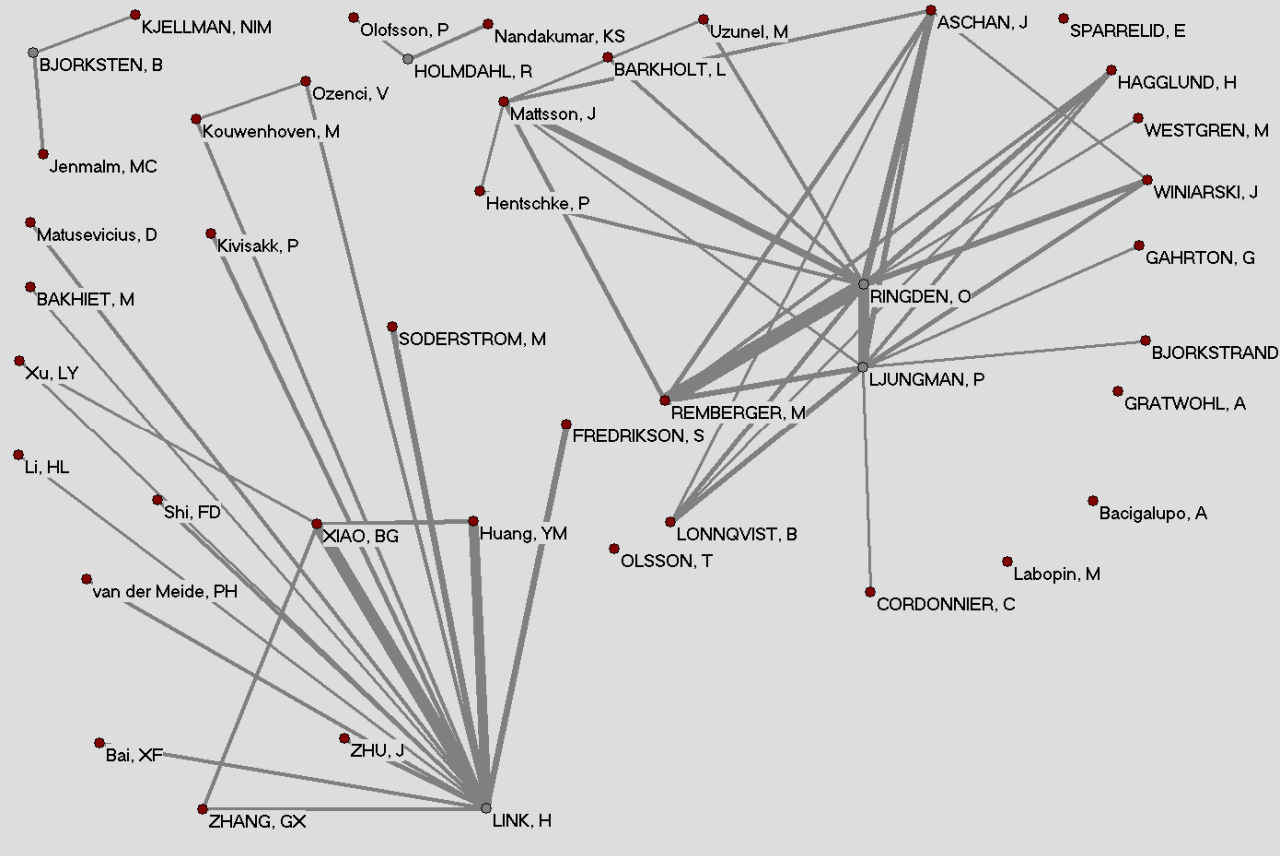

Figure 19. A co-publication map of the five most prolific Swedish journal article authors that have published in immunology journals 1995-2005 (Ringden O, Ljungman P, Bjorksten B, Holmdahl H and Link H). The selections of authors is the same as in Figure 17, that is each author included in the analysis has published at least 15 articles together with one or several of the selected Swedish authors. The connections shown in this map however, are at least 15 publications strong (suggesting a rather strong connection between the authors).

These three alternatives can be used alone or in any combination.

Subject interrelatedness

If data contains a subject classification of the publications, statistical methods can be used to analyse for example:

- Which subjects a researcher specifies in and how these connect him to other authors.
- Research areas that are connected by a certain number of publications.
- If research areas that show up frequently in the publications of a department are organized in one large departmental research network or several smaller ones.

How is bibliometrics used at Karolinska Institutet?

The possible use of data from Thomson ISI to analyse research publications from Karolinska Institutet has been studied since 2002. In 2005 a bibliometric pilot study was performed in cooperation with the Swedish Research Council.

In late 2005 the management of Karolinska Institutet decided that bibliometrics will be used as a tool in the ambition to become the leading medical university in Europe by 2010. This led to the formation of the project Karolinska Institutet Bibliometrics in the beginning of 2006.

The project Karolinska Institutet Bibliometrics

The project uses data about international scientific publications from the Thomson citation indices 1995-2008 to build a system capable of analysing Karolinska Institutet research publications and compare these to international publications. In the beginning of 2006, 10 million publication records from 1995-2005 were bought and loaded into a database. Additional records are downloaded weekly into the database as they are delivered from Thomson Scientific and the current Karolinska Institutet bibliometric database is expected to be as updated as the original Thomson databases.

Karolinska Institutet researchers are regularly asked to log in to an internal web site and verify their publications in the bibliometric database, something that increases the quality of both data and bibliometric analyses of Karolinska Institutet publications.

The primary goal is to supply the Karolinska Institutet management with analyses on Karolinska Institutet research publications as part of the institute's quality management. Secondary goals are to give researchers and employees working at the institute the opportunity to order analyses and to spread information about Karolinska Institutet publications and publication patterns.

A bibliometry-oriented customer service

Apart from analyses made for the Karolinska Institutet management, researchers and other employees of Karolinska Institutet have the possibility to order bibliometric analyses from the University Library at an individual cost.

References

[Anonymous]. (2005). Quantitative Indicators for Research Assessment - A Literature Review (Review No. 05/1). Canberra: Research Evaluation and Policy Project, Research School of Social Sciences, Australian National University

Chen, P., Xie, H., Maslov, S., & Redner, S. (2006). Finding Scientific Gems with Google: arXiv.

Debruin, R. E., & Moed, H. F. (1990). the Unification of addresses in scientific publications. In L. Egghe & R. Rousseau (Eds.), Informetrics 89/90 : selection of papers submitted for the second International Conference on Bibliometrics, Scientometrics and Informetrics, London, Ontario, Canada, 5-7 July 1989 (pp. 65-). Amsterdam: Elsevier.

Garfield, E. (1979). Citation indexing : its theory and application in science, technology, and humanities. New York: Wiley.

Glänzel, W. (2003). Bibliometrics as a research field: A course on theory and application of bibliometric indicators.

Lundberg, J. (2006). Lifting the Crown. Accepted for Journal of informetrics.

Lundberg, J., Fransson, A., Brommels, M., Skar, J., & Lundkvist, I. (2005). Is it better or just the same? Article identification strategies impact bibliometric assessments. Scientometrics, 66(1), 183-197.

Medical Subject Headings. (2006, 2006-08-25). Retrieved Sept 27th, 2006, from http://www.nlm.nih.gov/mesh/

Moed, H. F. (2002). The impact-factors debate: The ISI's uses and limits. Nature, 415(6873), 731-732.

Moed, H. F. (2005). Citation analysis in research evaluation. Dordrecht: Springer.

Oxford English dictionary (2000). Retrieved 060927, from http://dictionary.oed.com/

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics. Journal of Documentation, 25(4), 348?

PubMed Overview (2006, June 30, 2006). Retrieved Sept 20th, 2006, from http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html

Roth, D. L. (2005). The emergence of competitors to the Science Citation Index and the Web of Science. Current Science, 89(9), 1531-1536.

SciFinder - Content at a Glance - Patent and Journal references (2006). Retrieved Sept 20th, 2006, from http://www.cas.org/SCIFINDER/capluscontent.html

Scopus Info - Scopus in Detail - What does it cover? (2006). Retrieved 2006-09-20, 2006, from http://www.info.scopus.com/detail/what/

Seglen, P. O. (1992). The Skewness of Science. Journal of the American Society for Information Science, 43(9), 628-638.

Sentinel Event Glossary of Terms. (2006). Retrieved Sept 27th, 2006, from http://www.jointcommission.org/SentinelEvents/se_glossary.htm

Thomson Scientific: Citation Products. (2006). Retrieved April 6th, 2006, from http://scientific.thomson.com/products/categories/citation/

Ulrichsweb.com. Advanced Search. Retrieved April 6th, 2006, from http://www.ulrichsweb.com/ulrichsweb/Search/advancedSearch.asp?

Van Leeuwen, T. N. (2004). Second generation bibliometric indicators: the improvement of existing and development of new bibliometric indicators for research and journal performance assessment procedures. Leiden ,.

van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & van Raan, A. F. J. (2003). Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. Scientometrics, 57(2), 257-280.

van Raan, A. F. J. (2000). The Pandora's Box of Citation Analysis: Measuring Scientific Excellence - the last Evil? In B. Cronin & H. B. Atkins (Eds.), The web of knowledge : a festschrift in honor of Eugene Garfield (pp. 301-320). Medford, N.J.: Information Today.

What's the Difference Between MEDLINE and PubMed? Fact Sheet. (2006, 21 December 2004). Retrieved Sept 20th, 2006, from http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html