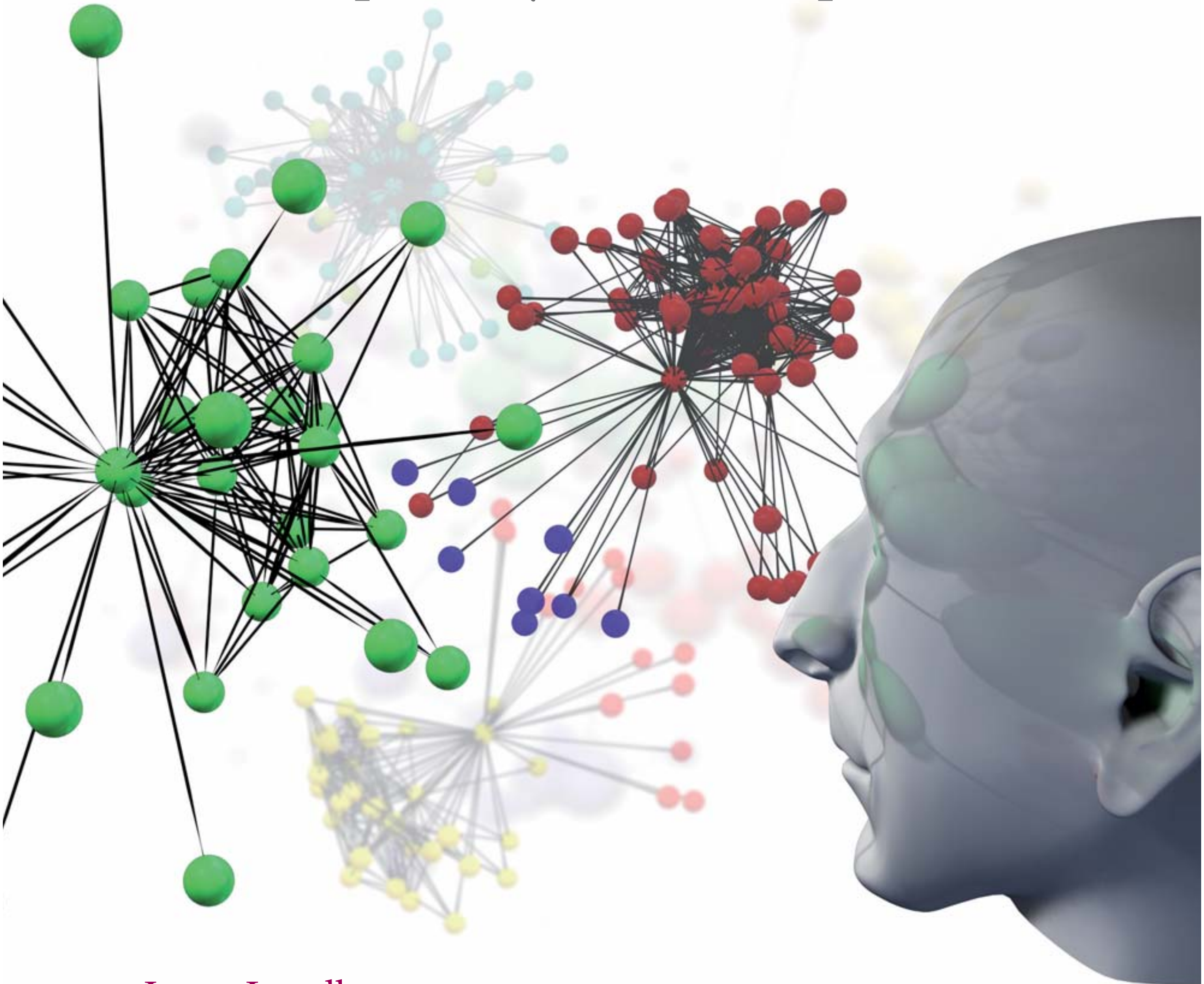
Thesis for doctoral degree (Ph.D.)
2006

Bibliometrics as a research assessment tool - impact beyond the impact factor



Jonas Lundberg

KAROLINSKA INSTITUTET ANNO 1810
Karolinska Institutet

Medical Management Centre
Department of Learning, Informatics, Management and Ethics
Karolinska Institutet, Stockholm, Sweden

Bibliometrics as a research assessment tool - impact beyond the impact factor

Jonas Lundberg

Stockholm 2006

All previously published papers were reproduced with permission from the publisher.

ISBN 91-7140-965-3

To Karolin

Abstract

Introduction: While bibliometrics is frequently used as a tool for research assessment and is thought of as objective, quantitative and unobtrusive, the usability of bibliometric indicators has seldom been properly assessed and their value therefore been questioned. The aim of this thesis is to explore and develop the utility of bibliometrics as a research assessment tool. This is done through four studies that address the validity of bibliometrics as a research assessment tool. The issues that are further investigated are related with field delineation (*Study I*), collaboration (*II*) and research performance (*I/III/IV*).

Materials and methods: The thesis is primarily based on data from the citation indices (CI) produced by Thomson Scientific. In the different studies bibliometric indicators are calculated based on 24 223 (*I*), 62 104 (*II*), 6 142 055 (*III*) and 596 (*IV*) publications. To assess the validity of bibliometric indicators calculated using the CI, these are combined with data from PubMed (*I*) and compared with data from manual assessments (*I*), financial data (*II*) and a Swedish system for identification and early assessment of new methods in health care (*IV*). Three new indicators are developed based on theoretical reasoning (*III*).

Results: Frequently used bibliometric methods do not allow valid assessment of the development of research areas (*I*) or collaboration between academia and industry (*II*). There are also flaws associated with the current state-of-the-art performance indicator (*III*). At the same time, there are bibliometric methods available that can be used for identifying research areas (*I*) and there are certain types of collaboration between academia and industry that accurately can be described using bibliometric indicators (*II*). New performance indicators, which assign equal weight to all publications and control for the skewed and differing distribution of citations over publications, have been developed (*III*). Publications related to health care technologies that are deemed as very promising by clinical experts receive high scores using these indicators (*IV*).

Discussion: The results of this thesis show that uncritical assessments of research areas based on rudimentary article identification strategies or collaboration analyses based solely on co-authorship data may be misleading and thus provide incorrect information for decision-making. At the same time, a correct use of refined bibliometric indicators may provide valuable background information for decision makers.

Keywords: Bibliometrics, research assessment, tool, collaboration, citation, indicator.

List of Publications

I. **Lundberg J,** Fransson A, Brommels M, Skår J, Lundkvist I. Is it better or just the same? Article identification strategies impact bibliometric assessments. *Scientometrics*. 2006;66(1):183-197.

II. **Lundberg J,** Tomson G, Lundkvist I, Skår J, Brommels M. Collaboration uncovered: Exploring the adequacy to measure university-industry collaboration through co-authorship and funding. *Scientometrics*. 2006;69(3)

III. **Lundberg J.** Lifting the crown – citation *z*-score. *Journal of Informetrics*. Accepted for publication

IV. **Lundberg J,** Brommels M, Skår J, Tomson G. Breakthroughs and bibliometrics: The sources of the Swedish rapid evaluation system of technologies in health care. *Submitted*

Contents

List of abbreviations

CI	The citation indices produced by Thomson Scientific. These include the Science Citation Index (SCI), Social Science Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI).
CWTS	Centre for Science and Technology Studies at Leiden University, the Netherlands
MeSH	Medical Subject Headings
P	Total number of publications
C	Total number of citations
$\bar{c}$	Average number of citations per publication
μ_f	Field reference value (field citation score) for articles of the same type, age and in the same field of research
$\bar{\mu}_f$	Mean field reference value (mean field citation score)
$\left[\bar{c}\right]_f$	CWTS field normalized citation score (crown indicator)
$\bar{c}_f$	Item oriented field normalized citation score average
Σc_f	Total field normalized citation score
$\bar{c}_{fz[\ln]}$	Item oriented field normalized logarithm-based citation *z*-score average (citation *z*-score)
I_{ISI}	ISI journal impact factor

1 Prelude

The work with this thesis can be compared to a long journey. It did not start from the shoulders of giants. I did not even know who or where the giants were. After some struggling I found two giants in the bibliometric countryside. On the way up trying to reach their shoulders I realized that the two giants would not provide the same view. One of them told me that bibliometrics was objective, quantitative and unobtrusive. The method was based on one of the most important outputs of the research process and did not require additional work from the researchers that would steal more time away from their already limited time for research. Bibliometric analysis could provide answers to questions asked by research managers across the globe: Who knows what? Who does what? Who knows who? It would enable us to assess the development over time and benchmark ourselves against our international competitors. The other giant told me that bibliometric methods reduced the endless frontier called science and human behaviour such as collaboration, to mere numbers. Being based on scientific publications it excluded other important outputs of the research process such as the training of PhD students and the development of new products. And what about implementation? Does the responsibility of a university end with the publication of results in scientific journals? Not even within its supposed stronghold, objective assessment of scientific publications, did bibliometrics seem applicable. The indicators were biased in favour of basic science in the western world. I decided not to choose path and climbed up to the top, placing one foot on the shoulder of each giant…

2 Introduction

2.1 The concept

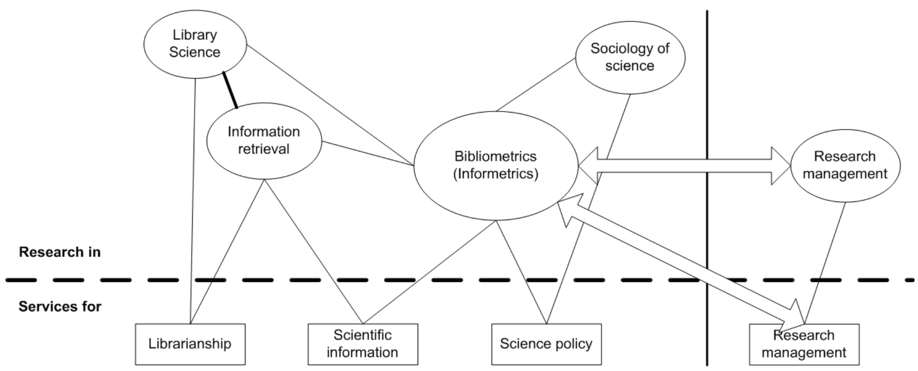
Bibliometrics as a term was introduced by Pritchard in 1969.[1] At the same time Nalimov and Mulchenko introduced the term scientometrics.[2] Pritchard defined bibliometrics as "*the application of mathematical and statistical methods to books and other media of communication*", while Nalimov and Mulchenko defined scientometrics as "*the application of those quantitative methods which are dealing with the analysis of science viewed as an information process*". Today the two terms are used almost as synonyms. The more general term informetrics is used when talking about quantitative analysis of various communication processes in science, i.e. not only scientific articles or books, but also electronic media and the aspects of information retrieval.[3]

The history of the field goes back further than the definition of the term. The earliest bibliometric studies took place in the 1920s when Alfred J Lotka investigated frequency distribution of scientific productivity (the number of people producing n papers is proportional to $1/n^2$) and Gross and Gross published on the subject of using citations as a basis for selecting what journals should be acquired by a university library.[4, 5] In the midst of the 20th century, important theoretical contributions for the development of the field were made by among others Derek deSolla Price, Robert Merton and Eugene Garfield.[6-11] One of the most important practical contributions was made by Eugene Garfield through his creation of the *Science Citation Index* in 1964.[12]

In the 1970s methods for (1) co-citation analysis for mapping scientific fields, (2) scientific collaboration, (3) growth of science, (4) citation patters for assessing scientific progress and (5) mobility in science were developed.[13-17] Francis Narin's development of the concept 'evaluative bibliometrics' (as compared to 'descriptive bibliometrics') was also important for the development of the field in the 70s.[18, 19] Evaluative bibliometrics concerns the application of bibliometrics with a focus on evaluation of science and more specifically on quality of scientific performance.[18] In 1978 the first journal with a specific focus on the field, *Scientometrics*, was founded by Tibor Braun. Other important journals for researchers within the field include *Research Policy*, *Journal of the American Society for Information Science* and *Research Evaluation*. A more recent addition to field is the *Journal of Informetrics*. In the 1980s (6) co-word techniques were developed and (7) the first comparisons between peer opinion and bibliometric indicators were made.[20-23] The first application of science indicators at the level of research groups was also made in the early 1980s.[24]

Today bibliometrics is closely related to research within the areas of library science, information retrieval and sociology of science (Figure 1). Research within in these areas are used as the basis for services for librarianship, scientific information and science policy.[3] Bibliometrics has also been used as a research management tool and a tool for research into research management.[25]

Figure 1 - Links of bibliometric with related fields and application services (modified from Glänzel[3])



2.2 Bibliometric theory

An important basic assumption within the field of bibliometrics is that scientific literature represent scientific activity.[7] A requirement for this assumption to hold is that researchers publish. Publishing papers has been a way to settle priority conflicts for a long time.[6, 8, 9] These priority conflicts occur since although some discoveries come by surprise, many more are more or less expected and a lot of people are working on them at the same time.[26] These conflicts have mainly been settled by claim staking rather than avoiding conflicts by sharing information. It has thus been suggested that researchers have a strong urge to write papers but a milder urge to read them.[6]

From a research management perspective (especially within the medical and natural sciences) the main focus of bibliometric studies are on publications in scientific journals. It is assumed that although there are other forms for communicating science, all important research findings are sooner or later reported in the serial literature.[27] This also implies that fields where internationally oriented scientific journals are not the main carrier of science communication, evaluative bibliometrics is less applicable.[28]

2.2.1 Reasons for giving citations

Garfield suggested that at least fifteen reasons exists for giving citations and Duncan et al. later suggested 26 different relations between cited and citing document (Table 1).[29, 30]

Table 1 – Reasons for giving citations / Relations between cited and citing documents

Garfield[29]	*Duncan et. al*[30]
1. Paying homage to pioneers	1. paying homage
2. Giving credit for related work (homage to peers)	2. background reading
3. Identifying methodology, equipment, etc.	3. historical;
4. Providing background reading	4. bibliographical leads
5. Correcting one's own work	5. narrative
6. Correcting the work of others	6. definition
7. Criticising previous work	7. clarification
8. Substantiating claims	8. illustration
9. Alerting to forthcoming work	9. example
10. Providing leads to poorly disseminated, poorly indexed, or uncited work	10. experimental detail
11. Authenticating data and classes of fact – physical constants, etc	11. theory
12. Identifying original publications in which an idea or concept was discussed	12. data
13. Identifying original publications or other work describing an eponymic concept or term	13. methodology
14. Disclaiming work or ideas of others (negative claims)	14. description
15. Disputing priority claims of others (negative homage)	15. current concerns
	16. development of ideas
	17. disputing
	18. criticism
	19. corroboration
	20. disclaiming
	21. substantiation
	22. similar research
	23. contradictory research
	24. further detail
	25. same paper
	26. statistics

More recently Shadish et al. and Case and Higgins have conducted large-scale surveys investigating motives for citation.[31, 32] Shadish showed that highly-cited articles are more likely to be thought of as exemplars.[31] Case and Higgins' asked researchers on what was most important for the decision to cite another study. The statement "*this reference reviews prior work in this area*" was the most commonly selected statement. It was followed by "*this reference is a 'concept marker'*" and "*this reference documents the source of a method or design feature*".[32] Neither Shadish nor Case found many 'negotional' types of citations (5%-9% were at least partly negative). These results are a bit higher than Chubin and Moitra and Oppenheim and Renn's findings of 2%-4% partially negotional citations, but lower than the 14% identified by Moravcsik and Murugesan.[33-35] The negotional citations most often refers to poor research and this gives rise to a J-shaped relation between research quality and citedness. Nicolaisen showed this J-shaped relation between quality of sociology monographs and their citedness, i.e. poor research was more highly cited than mediocre or ordinary research.[36] An early notion of this citing behaviour was made by Gilbert in 1977. "*The participants in a mature field will share a belief that some published work is important and correct, some other work is trivial, perhaps some is erroneous, and much is irrelevant to their current interests. Hence, authors preparing papers will tend to cite the 'important and correct' papers, may cite 'erroneous' papers in order to challenge them, and will avoid citing the 'trivial' and 'irrelevant' ones.*"[37] The existence of various motives for citing other documents seems obvious, but in the application of citation indicators it is assumed that given sufficiently large numbers, different motives for citing articles evens out.[38]

There is an ongoing debate on whether researchers give references (1) to register intellectual property and peer recognition and that counting citations thus measures intellectual influence, *or* (2) primarily for rhetorical purposes, i.e. to advance the interests of a researcher, defend his or hers claims against attack, convince others and thus gain a dominant position in their scientific community.[39-42] The former has been called a 'normative' or 'Mertonian' view, while the later has been called a 'social constructivistic' or 'Latourian' view.[39] The two viewpoints could either be considered as two extreme positions in the debate on citation theory, or as two perspectives representing two different dimensions of the scientific process. The basic assumptions guiding the sub studies of this thesis are closer to the ideas of the Mertonian dimension than the Latourian. Notions such as "*if one's work is not being noticed and used by others in the system of science, doubts of its value are apt to rise*" are supported.[43] It should be noted that this thesis is not concerned directly with referencing behaviour, but rather deal with various bibliometric indicators of which some are based on citations. A linkage to citation theory makes such indicators more interpretable, but the indicators are intended to be tested against empirical data. Different indicators may be used for answering different hypotheses. Some citation measures might be indicators of 'research quality' others are more likely to be indicators of 'influence' or 'impact' on the scientific community. The author therefore only partially agree with Cole and Coles notion that citation measures should be conceived as socially defined quality.[44] Other citation measures might measure innovativeness, self-centeredness, degree of internationalisation etc. Taking a broader perspective on bibliometrics, indicators exists that are supposed to test hypotheses on for example collaboration between academia and industry or development of research areas.

2.3 Bibliometric indicators

An 'indicator' can be defined as "*that which serves to indicate or give a suggestion of something; an indication of*".[45] The indicator in it self does not represent an 'absolute truth' and the definition also signals that the results describe a complex reality that can not measured merely by a single statistics or numbers. Hold stated "*to start with, by definition the measurement process extracts just one aspect of the object it is measuring and assigns a number to that aspect. But things in the real world are not described by a single aspect.*"[46] Another definition of indicator is: "*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.*"[47] This is similar to how the word is used in bibliometrics. When discussing what constitutes an indicator, it is useful to also consider the term data. The number of citations of one publication at a certain time is data, while the result of a mathematical operation with data to address some assumption is an indicator. As van Raan concluded "*indicators must be problem driven, otherwise they are useless.*"[48]

Science and technology indicators originated in the United States where the first Science Indicators report was published in 1973. The early indicators where focused on input and throughput indicators but lacked output and impact indicators. Even though publication based indicators are often referred to by science policy makers, there have been no accord of cases when policy decisions were solely based on publication (or patent) indicators. Instead the indicators have been used to justify decisions or as support in disputes.[49]

2.3.1 Performance indicators

Performance indicators are generally concerned with input, process or output measures. Bibliometric performance indicators can be considered as output indicators from the research process, while for example funding indicators are input indicators. Most bibliometric performance indicators are intended to assess performance along two dimensions: quality and quantity. Quantity indicators often involve the total number of publications that a unit has produced, while quality indicators concern various kinds of average citation rates. Indicators like total number of (field normalized) citations combines the two dimensions into something that gives an indication of the total impact that a unit has had on the scientific community. With regards to bibliometric indicators the focus of this thesis is on normalized citation indicators. The ISI impact factor and the h-index, two of the most heavily discussed bibliometric indicators, are also briefly discussed. An extensive literature review of quantitative indicators for research assessment has been carried out within the ARC Linkage Project in Australia and can be accessed on-line.[50] The literature review contains detailed descriptions and definitions of most bibliometric indicators used world wide.

2.3.1.1 The ISI Journal Impact Factor

The impact factor is a measure based on the average number of citations received the last year by articles published in a journal during the two preceding years (Box 1).[51, 52] The indicator is thus focused on the assessment of journals. One could imagine transferring the calculations to other assessment units (e.g. universities or research groups), where the publications included in the analysis are the ones of the unit instead of those in a journal.

Box 1 - Calculation of the ISI Journal Impact Factor[50]

C_x = Cites in year x to articles published in the two preceding years

P_{x-1-2} = number of articles published in years x-1 and x-2

$I_{ISIx} = C_x / P_{x-1-2}$ = Impact factor in year x

The strength of the indicator includes it's independence of the size of the journal, that its comprehensible, stable and seemingly reproducible (it has been shown to be quite difficult to reproduce).[3] At the same time several problems associated with the impact factors have been identified[40, 53-60]:

- Inaccurate definition of citable documents
- Lack of differentiation between publication types
- Short citation window
- Not comparable between fields
- Hides variation in article citation rates that results from a very skewed distribution

The greatest problem with the indicator though, is not its definition or calculation, but rather it's utility in evaluations of scientists, a matter for which it was not designed. The use of indicators outside their context is a general problem when discussing their validity. As Hand has stated "*one might even go so far as to say that the measurements or measurement procedures themselves are not really the subject of the validation at all. Instead it is the use made of the measurement which may or may not be valid. An IQ test will not be valid as a measure of social anxiety, for example.*"[46] The validity of using the impact factor for ranking journals is thus higher than the validity of using impact factors for ranking researchers. It has been concluded that "*impact factors in their undifferentiated form are outdated and should not be used as measures in any evaluative context at all. Yet, they are probably the most popular bibliometric measure of all.*"[38]

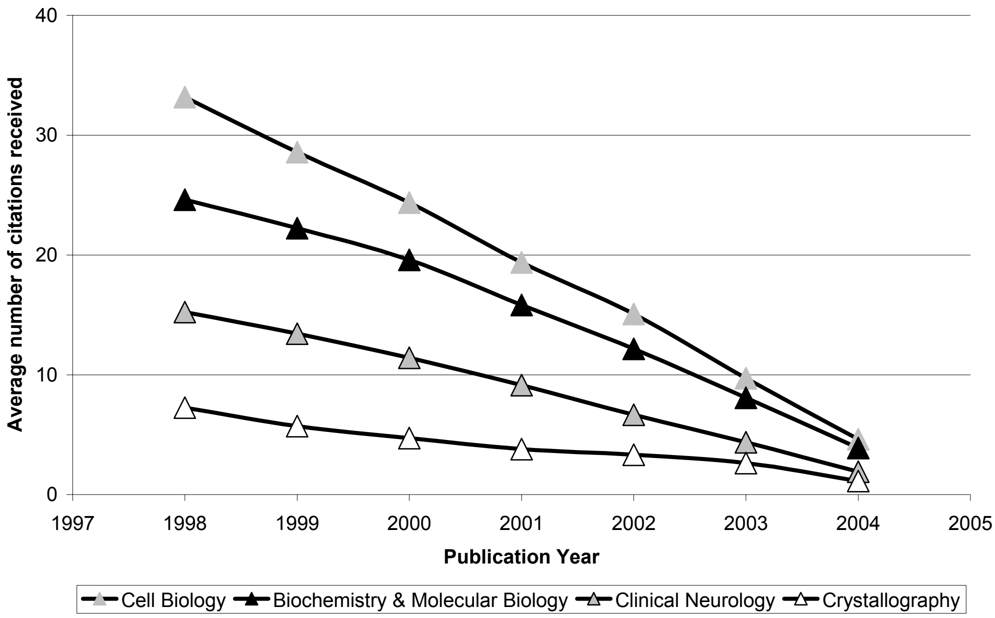
2.3.1.2 h-index

The *h*-index was suggested in 2005 by Hirsch.[61] The *h*-index is defined as "*A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.*"[62] A study by Bornmann and Daniel supported that the index might be useful. In their study the *h*-index was compared with peer review of applications for long-term fellowships to post-doctoral researchers. The authors concluded "*For every year, the h-indices of approved applicants are on average higher than those of rejected applicants*"[63] Even though the mean was higher for the approved group, the standard deviation was high which suggest that the *h*-index might not be suitable for assessing research on low aggregation levels. Another study by van Raan also partially supported the h-index by stating that it discriminated well between research groups rated excellent or good on the one side and less good (satisfactory) on the other. The index discriminated less well between groups rated good and excellent.[64] van Raan also noted that the 'crown indictor' (discussed later) relates better to the peer judgement than the *h*-index.[64] The *h*-index does neither take time nor research field into account.

2.3.1.3 Normalized indicators

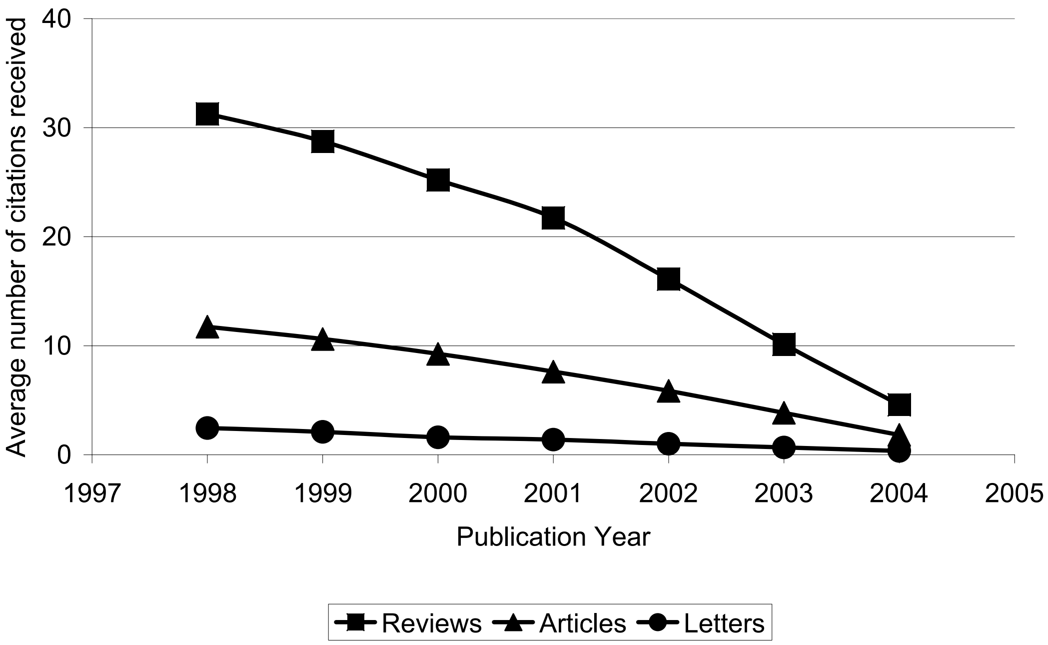
Beyond basic calculations like raw citation counts and the *h*-index, new indicators have been developed where citation rates are controlled for research field, publication year and document type. The reason for controlling for the first two factors is shown in Figure 2. As can be seen in the figure articles in different research fields and from different publication years have varying average citation rates. Articles in Cell Biology journals on average receive about five times as many citations as do articles in Crystallography journals. The difference between fields is quite consistent over time. The older a publication is, the more likely it is that it has been cited (regardless of field). As seen in the figure, articles in the displayed areas on average are cited between six (Biochemistry & Molecular Biology) and eight (Clinical Neurology) times more often if they were published in 1998 than if they were published six years later.

Figure 2 - Average citation rates for articles published in year 1998-2004 for articles in Cell Biology, Biochemistry & Molecular Biology, Clinical Neurology and Crystallography. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.



In Figure 3 an example is shown of why it is important to control for document type. Publications classified as *Articles* receive about two fifths as many citations as publications classified as *Reviews* (35%-40% during the year 1998-2004). *Letters* in turn receive one fifth as many citations as *Articles* (17%-21% during the years 1998-2004). In summary, there are obvious differences in average citation rates for publications of different types, of different age, published in journals within different fields.

Figure 3 - Average citation rates for articles, letters and reviews in CI published in 1998-2004. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.



The first suggestions on how to control for these factors and calculate 'normalized' citation rates were made in the 1980s.[65,66] In general, all proposed normalization methods are computed by dividing the actual number of received citations for a group of publications with the number of citations that could be expected for similar publications. Currently the state-of-the-art indicator is the so called 'crown indicator', developed at the Centre for Science and Technology Studies (CWTS) at Leiden University [67]. The indicator is calculated by dividing the average number of received citations for a group of publications with the average number that could be expected for publications of the same type, from the same year, published in journals within the same field.

2.3.1.3.1 Control for research area

It could be discussed whether to control for research field or to consider research fields with high average citation rate as being of higher quality than research fields with low average citation rate. One might argue that research performed within fields such as cell biology or biochemistry (with high average citation rates) on average is of higher quality than research within the fields of nursing or clinical neurology. One should then remember that the clinical neurologists easily could increase their average citation rate (and thus their 'quality' as measured by raw average citation rate) by starting to write longer reference lists. This would of course not increase the quality of research in clinical neurology.

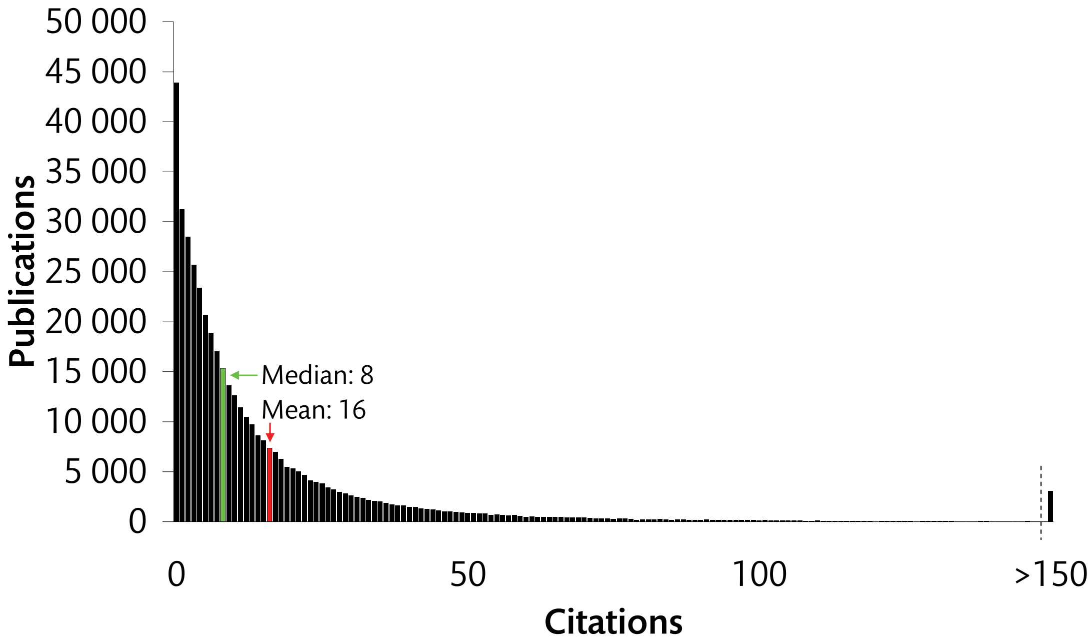
2.3.1.3.2 Control for document type

Another question is whether to control for document type. That review and 'normal' journal articles differ in average citation rates as well as in type of research is uncontroversial. A more contentious issue is whether to control for the document type letter, as these publications simply could be seen as short normal journal articles. Their average citation rate is however much lower than the citation rate of normal articles. In the suggested indicators, letters and normal articles are separated in the normalization procedure and the choice of which document types to include is made when calculating the final indicator of a researcher, group or department. One could simply choose to omit letters from the calculation of the final indicator values or display indicators for each publication type separately.

2.3.1.3.3 Current state-of-the-art bibliometric performance indicator

It can be argued that the current state-of-the-art bibliometric performance indicator (the CWTS crown indicator) has flaws. The first is that citation rates are not normalized on the level of individual publications, but on a higher aggregation level where the average citation rate of a researcher, group or department is compared to the average citation rate of the fields in which the researcher or group has published. This way of calculating gives more weight to older publications (particularly reviews), published in fields with dense citation traffic. Another problem is that the distribution of citations over publications differs between 'normalization groups' (publications of a specific type, within a specific research field, published a specific year). Therefore not only does the average citation rate differ, but also the standard deviation. Further is the distribution of citations over publications is highly skewed. The skewed distribution has been seen for fields and journals, as well as for individual scientists.[68] An example for life science articles published in 2000 is shown in Figure 4. Ten percent of the articles are still uncited six years after publications (including self citations). Sixty percent have received ten or less citations. At the same time, ten percent of the publications have received fifty percent of the citations, with some articles receiving more than 1000 citations. As with all positively skewed distributions, the mean is higher than the median. In this example the mean (16 citations) is twice as high as the median (8 citations).

Figure 4 – Distribution of citations over publications for 424 480 articles within 57 life science areas published in 2000 and citation received 2000-2006 (includes self-citations). 3052 publications (0.7%) have received more than 150 citations and are aggregated in the graph as '>150'.



2.3.2 Structural indicators

There are numerous 'structural' indicators that can be calculated. These include the character (basic or applied) of a unit's research, 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 units research field, or of their co-authors and the co-authors affiliations (organizations, countries etc).[50]

2.3.3 Collaboration indicators

A research collaboration can be defined as "*a system of research activities by several actors related in a functional way and coordinated to attain a research goal corresponding with these actors' research goals or interests.*"[69] This definition of research collaboration has been further divided into six categories by Laudel[69]: 1) Collaborations involving a division of labour; 2) service collaboration, 3) transmission of know-how; 4) provision of access to research equipment; 5) trusted assessorship; and 6) mutual stimulation. Laudel showed that it is collaboration involving a division of labour that primarily results in co-authored publications.[69] Laudel's definition is narrower than what was suggested by Katz and Martin, who suggested that such a definition is impossible since research collaboration has 'fuzzy' borders and potential borders are drawn as a matter of social convention.[70] They instead provided a checklist for how to distinguish 'collaborators' from other researchers. The criteria for who should be considered as collaborator included for example people working together on a research project for a large part of its duration, people appearing on the original research proposal or people responsible for main or key elements of the research. Katz and Martin further suggested that those who make only minor contributions to the project or are not 'proper researchers' (e.g. technicians) normally should not be seen as collaborators.

Analysing co-authored publications has become the standard way of measuring research collaborations.[6, 69-79] Studies of publications co-authored between organizations, i.e. analyses of multiple affiliation addresses given in journal articles, are used to assess links between organizations and for carrying out quantitative studies of collaboration patterns.[80] Co-authored publications grant access to an often informal network, can be viewed as successful scientific collaboration and indicate diffusion of knowledge and skills.[75, 81-84] Other advantages of assessing research collaborations through co-authorships are that the indicator is quantifiable and invariant, the measurement does not effect the collaboration process and the analysis is relatively inexpensive.[70, 85] Based on these advantages, co-authorship analysis has become one of the most frequently used indicators in bibliometrics.

At the same time, bibliometric researchers have advised that co-authorship based indicators should be handled with care as a source of evidence on actual scientific collaboration.[69, 70, 72, 75] Katz and Martin noted that few authors have examined the adequacy of measuring collaboration through co-authorship and that there are few systematic studies assessing the validity of the co-authorship indicator.[70] Katz and Martin also cautioned researchers from using co-publication as a sole indicator of collaboration since there are many cases when collaboration will not result in a co-authored paper.[70] University researchers might for example publish results of joint efforts without mentioning the direct involvement of industrial researchers.[75] These collaborations would not be identified using a bibliometric approach. It can also be that publications appear to be inter-institutionally co-authored when no collaboration has taken place. This could occur when a researcher has moved from a university to industry and in his/her publication lists both the prior and current affiliation. Melin and Olsson concluded that even though there is a risk of omitting collaborations in co-authorship studies, the tendency is rather that collaboration will be overrepresented when studying co-authorships.[72]

2.4 Methodological considerations in bibliometric assessments

There a number of methodological considerations to deal with when carrying out bibliometric assessments. The notion that indicators extract one single aspect of a complex reality[46] together with the fact that data sources are imperfect and that there is a choice of what indicators to use in assessments all underline the importance of critically discussing issues related to bibliometric methods.

2.4.1 Field delineation

Bibliometric analysis has been performed within and across national, institutional and individual levels in the evaluation of research areas.[86-94] An early step in bibliometric analysis is the identification of relevant scientific articles. For analysis of research areas this step concerns the identification of articles belonging to the area under study. Co-citation and co-word techniques have been developed and also been used in combination with citation analysis in bibliometric evaluations.[13, 20-22, 95] Still researchers and analysts often use either journal classifications or keywords as the basis for their article identification strategy.[86-89, 91-94, 96-98] The most frequently used journal categories are the ISI Subject Categories. A general concern with using this method is that it has been shown that publications within one field are often published in journals that are categorised as belonging to another field.[99, 100] A related matter concerns articles published in multidisciplinary journals such as Science or Nature. Keyword searches may be based on free text searches in titles and abstracts, or based on searches of keywords assigned to articles by experts. Within the medical field the controlled vocabulary named Medical Subject Headings (MeSH) is the most frequently used expert assigned keyword system.[101] Another frequently applied strategy to identify relevant articles is the use of names of researchers within the specific area.

2.4.2 Quality or quantity?

Research assessments most often need to consider quantity as well as quality. Calculation of composite indicators has been suggested by for example Lindsey.[102] In this thesis a Σc_f is suggested that gives an indication of the total international impact that an assessed group, organisation or country has had (*III*). It should be noted that extreme 'quality' indicator values (both low and high) are to be expected for units with a low number of publications. This is of importance when using the suggested indicators for comparing units of different size (for example small and large research groups within a university).

2.4.3 Aggregation level

The level of aggregation should be based on what one intends to measure. If the aim is to evaluate universities this ought to be the aggregation level chosen for the bibliometric assessment. A more difficult question could be whether to assess research within a university on the level of individual researchers or that of research groups. A research group contains more than one researcher and thus provides a larger base for statistical analyses. In general a research group produces 10 to 20 publications per year which have been considered as a sufficient number.[103] Furthermore, discrepancies exist in publication behaviour where some group leaders tend to require co-authorship on all publications within their group, while other group leaders do not. It could thus be considered as more fair to analyse senior researchers on the level of their groups. At the same time, it might be the case that research group leaders under such conditions would 'free-ride' on the work of independent scientists within their group. The decision of aggregation level is important since it influences the results of the bibliometric assessments. The decision might also influence publication practice within the studied entity. If the choice is made to conduct analyses on the individual level, research group leaders have incentive to require co-authorship from their group members.

2.4.4 Time frame

In general a time frame of five to ten years is used in bibliometric analyses, but this should be decided based on the objectives of the study.[104] The impact of the time frame on the validity or predictive value of bibliometric indicators has not been thoroughly assessed.

2.4.5 Retrospective or prospective focus

Depending on the question the bibliometric assessment is intended to assist in answering a decision must be made on whether to use a 'retrospective' or 'prospective' focus. This decision is illustrated by the example in Figure 5. In this example the aggregation level 'university' and the time frame 1995-2005 are chosen. There are three researchers (I-III) at the universities in the example. Researcher I resided at University A during the whole time period. Research II moved from University A to University B in 2000 while Researcher III moved from B to A in the same year. If we whish to assess university A, then what publications must we include? If we are interested in the past performance of the unit of analysis, for example how many articles University A has published per $ spent on research during the time period we should include everything that all researchers have produced while being at the university. We would thus include all publications from Researcher I, publications between 1995 and 2000 for Research II and publications between 2000 and 2005 from Researcher III (Figure 6). The assessment in this case has a *retrospective* focus. If instead we are interested in predicting what a university will perform the next year or set out to assess the competence level of University A, we should disregard publications by Researcher II since this researcher no longer is affiliated with the university. Instead we should consider all publications by Researcher III, even though the publications 1995-2000 where not produced at University A (Figure 7). If this kind of focus is chosen it can be called *prospective*, or directed at the future. Thus, the choice of retrospective or prospective focus is not a matter of right or wrong but rather a decision to make dependent on what question the bibliometric assessment is intended to assist in answering.

Figure 5 - Illustration of the difference between retrospective and prospective focus of bibliometric assessments. When evaluating the two universities (A&B) what publications by researchers I-III should be included?

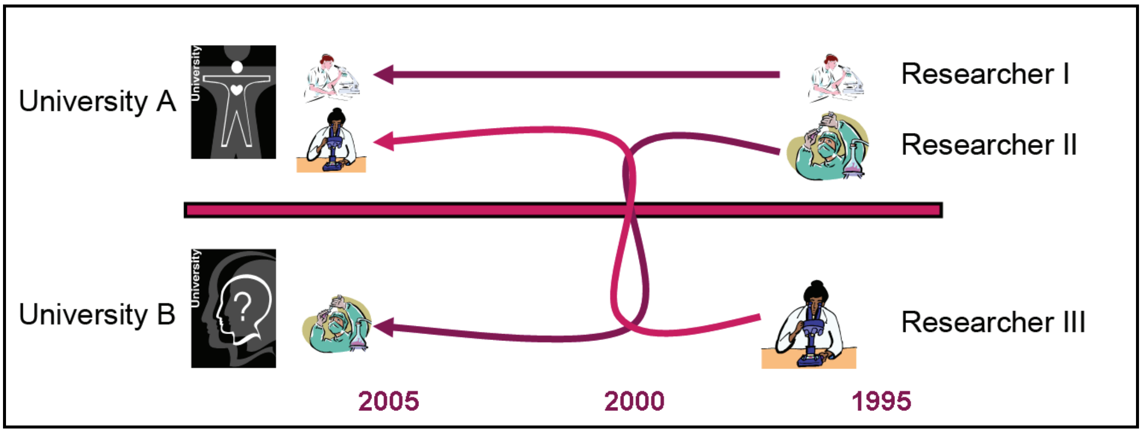


Figure 6 - Publications to include when using a retrospective focus (publications to exclude are shadowed)

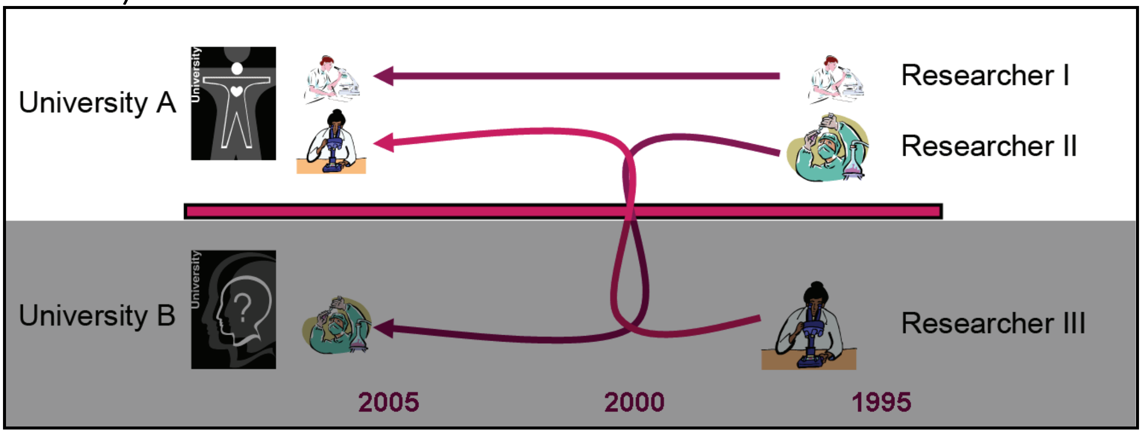
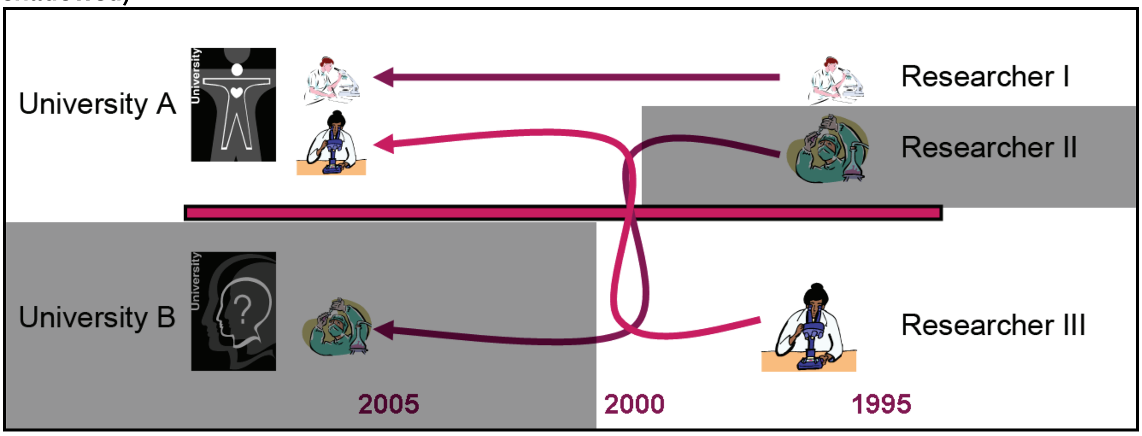


Figure 7 - Publications to include when using a prospective focus (publications to exclude are shadowed)



2.4.6 Counting scheme

Before carrying out a bibliometric assessment a decision is made on what counting scheme to use. There are three main counting schemes that are used within the bibliometric community: (1) fractional counting (2) first address count and (3) full counting (Box 2).[3] One could also imagine the use of weighting publications by where in the author list researchers appear. For example could first and last authors be given higher credit for a paper than researchers appearing somewhere in the middle. This would be based on the notion that in some research areas in certain countries (e.g. medical research in Sweden) it is a convention that the junior scientist who have carried out the main part of the work in a research project is first author, while the senior scientist responsible of for example funding of a project is put last in the sequence of author. Note that this is not the case for all research areas. Therefore it is only applicable in specific contexts. Exactly how different counting schemes affect the validity of different bibliometric indicators is poorly assessed, but there is currently a trend within the bibliometric research community towards an increased use of fractional counting.

Box 2 - Three common counting schemes[3]

1. *Fractional counting* – if *n* units (authors, institutions, countries, etc.) have contributed to the paper, each takes the value $1/n$ for the paper.
2. *First address count* – a paper is assigned to one unit only, on the basis of the first address in the address list of a paper as included in the database.
3. *Whole counting* (also called *full* or *integer counting*) – assigns a publication fully to each contributing unit.

An interesting case of fractional counting was seen in a report from the Swedish Research Council in 2006.[105] In the report was not only fractional counting conducted for calculating quantity indicators such as total publication volume, but also for more quality oriented indicators such as the ‘crown indicator’. It is not unreasonable to believe that high quality research involves a larger number collaborating organisations (or countries) than low quality research. A low quality article by a single author at a university would therefore have a higher impact on the assessment of a university than would a high quality publication that involves multiple universities. It is statistically appealing to use fractional counting, but its effect on the validity of bibliometric indicators needs further assessment. Both fractional and whole counting can have an impact on the behaviour of scientists. The use of whole counting is likely to stimulate collaboration, but also less positive behaviours such as honorary or false co-authorship.

2.4.7 Self-citations

Another important subject is self-citations. It has been shown that the share of self-citations is higher in co-authored papers and even more so in internationally co-authored papers.[106] Lately a decrease in self-citations has been noted. Between 1991 and 1999 the share of self-citations decreased from 29.1% to 26.7%.[107] Studies have shown that self citations do not significantly influence analysis results when you study a sufficiently large number of publications.[3] On the level of research groups it can have an effect and self-citations are therefore often removed (as much as possible) in modern bibliometric analyses.

2.4.8 Citation window

A citation window is the selected period during which citations to publications are counted. A two-year citation window for example means that for publications made in 1990, citations made during the year 1991 and 1992 are counted and for publications made in 1999, citation made during the years 2000 and 2001 are counted. A three-year citation window has been suggested to be suitable. This provides a compromise between areas where publications gets obsolete fast (e.g. life sciences and experimental physics) and slowly aging topics like theoretical physics.[108, 109] If normalized indicators are used, such as those suggested in this thesis, it is not necessary to use a citation window.

2.4.9 Data quality

Before proceeding with any bibliometric assessments an important (and often time consuming) step is required: validation and quality improvement of the data. One of the most important quality improvement steps is related to the existence of homonyms, synonyms and misspellings.

When working with assessments of individuals (and in some cases also organizations) it is important to note the existence of both homonyms and synonyms. By homonyms instances are meant where the same name represents two or more different things. This can happen when two different authors share the same name. By synonyms we mean instances when two different names represent the same thing. This can occur when scientists get married and change their last-name. Assessing researchers based on their current name would thus risk omitting important publications made prior to marriage.

Manual validation is also necessary to correct misspellings. Names can be (or become in the data handling process) misspelled on various levels, not least on an organizational level. Looking at Karolinska Institutet in the CI we have found among others the following spellings: 'KAROLINSKA INST' (correct!), 'KAROLINSA INST', 'KAROLINSK INST', 'KARDLINSKA INST', 'KAROLINSKI INST HOSP', 'KAROUNSKA INST', 'KAROLINSK INST & HOSP', 'ROYAL KAROLINSKA INST', 'HOSP KAROLINSKA INST', 'KAROLINKA INST & HOSP', 'KARMINSKA INST', 'KAROLINKSA INST', 'HISTOLOGEN KAROLINSKY INST', 'KAROLINSKI INST' etc.

The validation of bibliometric data is of great importance when conducting bibliometric analysis. Van Raan has given a clear statement on this matter, to which the author of this thesis subscribe.[110]

> *"Bibliometric analysis for evaluation purposes, or for studies of science in general, is and must be much more than just using what is readily available in databases. This is a crucial element of advanced bibliometric research, development and practical application. Scientists write their publications for communication and knowledge dissemination. Again, they are not responsible for the use of their publications for a completely different goal, namely evaluation of scientific performance. So 'bibliometric evaluators' have their own, specific responsibility that is different from and beyond that of an individual author, particularly as far the attribution of publications concerns, as precisely this attribution is one of the most crucial elements in the whole evaluation process! This is simply a matter of ethical conduct as an evaluator."*

2.4.10 Journal coverage

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 for analyses of areas within the humanities or social sciences other methods must be applied as well.[40]

2.4.11 Citation indicators and clinical research

There is a concern that there is a lack of recognition for applied disciplines and that there is a bias against clinical research in bibliometric assessments. [56, 111-113] Therefore it would be valuable to know whether publications of importance for improving the health system and its technologies are also highly cited intra-scientifically.

2.5 Bibliometrics and peer review

When assessing the validity of bibliometric indicators that are used for assessing research, the indicators are generally compared with peer review. As Roessner stated "*Peer review remains the backdrop against which all other types of research evaluation appear, and often the standard against which their validity is judged.*"[114] The correlation between peer review and citation counts was first identified by Clark in 1957.[115] In his study a panel of experts within psychology were asked to name those who contributed most to the field. The number of times a person was nominated was then correlated with various performance indicators. The highest correlation was seen for citation counts. The first comparisons between peer opinion and bibliometric indicators were made in the 1980s.[23] Many studies have since confirmed the correlation between bibliometrics and peer judgements.[40] Bibliometric indicators have been shown to be especially useful for discriminating between high and low quality, but less successful in discriminating within the large group of units of average research quality.[40, 64]

Most bibliometric researchers are eager to stress that quantitative analyses of research performance should enhance rather than replace peer review. The need for enhancing the peer review process stems from notions such as peer evaluation "…*diverge, they contradict each other, and they do not remain consistent over time.*"[38] The increased demand for research evaluations have also shown the limitations of traditional peer review, for example is the capacity for peer review, i.e. peers that are suitable for reviewing research, finite.[116]

The complementing use of peer review and bibliometrics has been explained by Weingart:

> "*The unique contribution of bibliometrics to the collective communication process in science and their greatest value to the scientific community itself as well as to policy makers and the public is in providing this 'greater picture'. However, bibliometric analysis and evaluation do not replace peer review for an obvious reason: The interpretation of these patterns, of unexpected contradictions to the common wisdom of the community or other irregularities must be left to the experts in the respective fields or at least assisted by them. Peer judgment must complement bibliometric analyses wherever necessary.*"[38]

3 Study rationale, aim and objectives

3.1 Study rationale

As described in the preceding section there are many issues associated with bibliometrics. The effects of these matters on the utility of bibliometric as a research assessment tool were in many cases unknown at the time of planning and carrying out this doctoral project. Not all concerns could be dealt with as part of this thesis project, but rather are a few questions selected to provide insight into different areas. The selected issues are related to research performance (*I/III/IV*), collaboration (*II*) and field delineation (*I*) as well as developing new (*III*) and validating existing indicators (*III/IV*). The knowledge gaps associated with these matters are summarized below.

Journal categories are frequently used for research area identification in bibliometric assessments. It has been shown that publications within one field are often published in journals that are categorised as belonging to another field.[99, 100] It is therefore of interest to know whether the choice of article identification strategy impact analysis of the development of research areas (*I*).

Also in the area of collaboration assessment the utility of bibliometrics has been questioned. Bibliometric researchers have advised that co-authorship based indicators should be handled with care as a source of evidence on actual scientific collaboration.[69, 70, 72, 75] Katz and Martin cautioned researchers from using co-publication as a sole indicator of collaboration since there are many cases when collaboration will not result in a co-authored paper. They also noted that few authors have examined the adequacy of measuring collaboration through co-authorship and that there are few systematic studies assessing the validity of the co-authorship indicator.[70] Additional studies of the validity of co-authorship is thus needed, as is the development (and use) of indicators that could complement co-authorship as an indicator of collaboration (*II*).

Another area in which bibliometric methods are applied is performance assessment. As described in the section on normalized indicators (p. 13) the current state-of-the-art indicator has flaws. These include that citation rates are not normalized on the level of individual publications and that the indicator neither takes the varying distribution of citations over publications between 'normalization groups' nor the highly skewed distribution of citations over publications into account. It could be argued that at least in theory, the validity of the indicator would be increased if the mentioned issues are dealt with (*III*).

Last is the issue of a potential lack of recognition for applied disciplines and the potential bias against clinical research in bibliometric assessments.[56, 111-113] This question needs further investigation to more clearly set the boundaries for bibliometric assessments. Therefore it would be valuable to know whether publications of importance for improving the health system and its technologies are also highly cited intra-scientifically. A way to explore this is to calculate bibliometric indicators for publications related to health care technologies that are deemed as very promising by clinical experts (*IV*). If bibliometric assessments are biased against clinical research it needs to be controlled for. On the other hand if they are not, it would increase the utility of bibliometrics as a research assessment tool.

3.2 Main aim

The aim of this thesis is to explore and develop the utility of bibliometrics as a research assessment tool.

3.3 Specific objectives

- Assess whether the choice of article identification strategy has an impact on bibliometric analyses of the development of research areas *(I)*.
- Assess how well university-industry collaborations can be identified and described using co-authorship data *(II)*.
- Introduce new bibliometric indicators that deal with flaws related to the current state-of-the-art indicator *(III)*.
- Assess whether publications of importance for improving the health system and its technologies are highly cited intra-scientifically *(IV)*.

4 Materials and methods

4.1 Data sources

The main data sources of this thesis are the citation indices produced by Thomson Scientific (formerly Institute of Scientific Information).[*] These include the Science Citation Index (SCI), Social Science Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI) and are jointly referred to as the Thomson Scientific Citation Indices (CI). The CI are the most widely used data sources for bibliometric analysis. Even though the CI have some important limitations, they still contain unique information that is relevant when conducting a bibliometric analysis. This includes *references* (enables citation analyses), *full addresses* (all addresses listed by the authors of the publications are listed in the database) and *multidisciplinarity* (CI covers most research fields with best coverage of the life sciences, natural sciences, mathematics and engineering). The CI are most readily available through their common on-line interface – Web of Science®. More complicated analyses, e.g. calculation of normalized citation indicators requires direct access to world-wide citation data. At Karolinska Institutet the development of a bibliometric analysis system (with CI data imported into a *MySQL* database) has taken place during the course of this thesis project. Therefore the first two studies (*I/II*) are based on data from the Web of Science®, while the last two (*III/IV*) are based on the bibliometric system at Karolinska Institutet. It could be noted, that normalized citation rates of individual publications (*III*) could be calculated and published by Thomson Scientific on their Web of Science®. Until this occurs the calculation of the suggested indicators requires direct access to world-wide citation data. In the four studies references to 24 223 (*I*), 62 104 (*II*), 6 142 055 (*III*) and 596 (*IV*) publications in CI served as the main basis for analysis.

In the first study of this thesis CI data was combined with data from PubMed. PubMed is a service of the National Library of Medicine (USA) that includes over 16 million citations from MEDLINE and life science journals for biomedical articles back to the 1950s.[117] PubMed contains value adding information, e.g. MeSH-terms (Medical Subject Headings) and is available free of charge. Limitations of using PubMed as a data source for bibliometric analyses include that before the year 2002 only corresponding author addresses are listed and that PubMed does not contain information on cited references. In study *I* 24 223 articles published during the years 1994-2003 by researchers affiliated with Karolinska Institutet, included in both PubMed and Web of Science®, were used to assess whether the choice of article identification strategy has an impact on bibliometric analyses of the development of research areas.

In the second study, financial information from the economic system of Karolinska Institutet was used as the source for a comparison of co-authorship data with industrial funding. Information on all 813 organizations (both companies and other organizations) that provided funding to Karolinska Institutet during the years 1993 and 2004 were collected. Companies working in R&D related areas were selected for the study. Companies within the areas of insurance, publishing or real estate and government-owned companies were excluded. Other inclusion criteria included that each company should have provided at least 50,000 SEK[†] in funding between 1993

---

[*] Certain data included herein is derived from the 1995-2005 Science Citation Index Expanded, Social Sciences Citation Index and, Arts & Humanities Citation Index Tagged Data prepared by the Thomson Scientific Inc (TS), Philadelphia, Pennsylvania, USA: ©Copyright Thomson Scientific Inc® 2006. All rights reserved.

[†] SEK = Swedish Kronor, 1 SEK ≈ 0,12 US$ (November 18th 2005)

and 2001. Companies were then grouped to the level 'current parent company'. In the end, 95 'current parent companies' were identified that met the inclusion criteria.

In the fourth study the 50 SBU Alerts published (or released in new versions) 2001-2004 were used to identify publications related to health care technologies that are deemed as very promising by clinical experts. Within the 50 SBU Alerts a total of 953 references were given. To calculate normalized citation scores for the references, these where matched with corresponding references in CI. After removal of references to items not included in CI and publications that were not of the types journal articles, reviews or letters, 596 references remained for further assessment. More details on the matching procedure are given in study *IV*.

4.2 Study Methods

4.2.1 Study I

Assessing the accuracy of a test most often concerns exploring its recall (or sensitivity) and precision (or positive predictive value). Recall is the proportion of cases in which the test is positive out of all cases where the outcome did occur. Precision in turn is the proportion of cases in which the test was positive and the outcome occurred out of all cases where the test was positive (Figure 8). From an information retrieval perspective recall can be defined as the proportion of relevant documents that are retrieved, out of all relevant documents available. Precision is the proportion of retrieved and relevant documents to all the documents retrieved.

Figure 8 - Calculation of recall and precision. Recall = a/c1; Precision = a/r1

	Outcome occurred (+)	Outcome did not occur (-)	*Totals*
Test positive (+)	a	b	*r1*
Test negative (−)	c	d	*r2*
Totals	*c1*	*c2*	*t*

When assessing the accuracy of article identification strategies recall is the proportion of relevant articles that are retrieved through a search strategy and precision the probability that an article that is identified through a search strategy actually belongs to the area. To know whether an article belongs to an area (if the 'outcome' occurred) a 'golden standard' is used. For article identification strategies actual article assessment performed by researchers within the areas under study (cancer/hearing) was used as golden standard. The calculations of confidence intervals for recall and precision were based on constant chi-square boundaries[118] using John C. Pelluzo's 2-way Contingency Table Analysis software.[119]

To assess whether the choice of article identification strategy has an impact on bibliometric analyses of the development of research areas, the development of two different areas (cancer and hearing) with regards to total number of publications per year, mean impact factor and average number of citations received two years post publication is described based on articles identified through three different strategies: journals, MeSH and author. The null-hypothesis was that the choice of strategy would not influence the assessments. If the different strategies result in varying assessments with regards to the development of average number of citations received, average impact factors and total number of publications the null-hypothesis should be rejected.

4.2.2 Study II

In the second study the level of funding from a company to a university was used as a comparative indicator of research collaboration (*II*). Other methods that would be suitable for comparative analysis are interviews, surveys or observation. Collaboration analysis using both co-authorship and surveys has been performed for example by Laudel.[69] The reasons for choosing funding as the comparative indicator are that funding is readily available (at least in a Swedish context), reliable (other researchers would yield the same results), possible to measure over time and quantifiable. Further, funding has to our knowledge not previously been used as a comparative indicator to co-authorship when measuring collaboration. When funding has been studied in relation to co-authorship, the focus has been on the correlation between number of authors involved in a paper and the level of financial support.[120] The choice of indicator is also based on the assumption that funding is a relevant indicator of collaboration between a university and industry and that companies primarily fund universities when engaging in research partnerships and not for public good.[75] Moreover, observation, interviews or surveys have been suggested as unsatisfactory for determining the precise nature and magnitude of collaboration.[85] Funding might be a 'golden standard' for identifying collaborations or at least a valuable comparison to co-authorship as collaboration indicator.

4.2.3 Study III

The third study is a critical assessment of the state-of-the-art bibliometric performance indicator. The development of three new indicators is based on theoretical reasoning, which is elaborated in the results section (page 33).

4.2.4 Study IV

The aim of the fourth study of this thesis was to assess whether publications of importance for improving the health system and its technologies are highly cited intra-scientifically. This was done by assessing the 'bibliometric impact' of publications used as sources in a Swedish system for identification and early assessment of new methods in health care developed by the Swedish Council on Technology Assessment in Health Care (SBU): *SBU Alert*. Normalized citation scores were calculated for all articles, letters and reviews published since 1995 that have been used as sources in SBU Alerts published (or released in new versions) 2001-2004 and are included in CI.

A general bibliometric analysis of the science base of SBU Alert was also conducted. The analysis include answering the questions of how old the used sources are, where the research was carried out, what countries that are over-/under represented, in what journals the research was published and to what subject categories the journals belong.

5 Results

5.1 Study I

The specific objective of the first study of this thesis was to assess whether the choice of article identification strategy has an impact on bibliometric analyses of the development of research areas.[121]

5.1.1 Using journal categories for article identification has low recall but high precision

Journal-based strategy had the lowest recall and, together with the MeSH strategy, the highest precision within the area of cancer research (Table 2). The MeSH strategy had the highest recall. For hearing research the recall was highest for the MeSH-strategy which also, together with the hearing-journal strategy (referred to as Journal_H in the study), had the highest precision.

Table 2 - Recall and precision (with 95% confidence intervals) for different search strategies – journals, authors, MeSH-terms – and research areas – cancer and hearing. Journal_H includes only specific hearing research journals. Journal_{ORL} includes all journals belonging to the otorhinolaryngology category.

Area	Strategy	Recall (95% CI)	Precision (95% CI)
Cancer	*Journal*	0.47 (0.43-0.50)	0.89 (0.81-0.94)
Cancer	*Author*	0.57 (0.51-0.63)	0.50 (0.44-0.55)
Cancer	*MeSH*	0.87 (0.83-0.89)	0.93 (0.89-0.96)
Hearing	$\textit{Journal}_H$	0.38 (0.27-0.38)	1.00 (0.70-1.00)
Hearing	$\textit{Journal}_{ORL}$	0.57 (0.39-0.72)	0.52 (0.36-0.66)
Hearing	*Author*	0.86 (0.70-0.94)	0.75 (0.61-0.83)
Hearing	*MeSH*	1.00 (0.91-1.00)	1.00 (0.91-1.00)

5.1.2 In what journals do cancer and hearing researchers publish?

The 370 cancer researchers included in the study published their results in at least 675 different journals during the period of 1994-2003. These journals belonged to 110 different ISI subject categories. 14 of the 25 most frequently used journals did not belong to the ISI subject category ‘oncology’ (Table 3). In total 73% of the articles published by cancer researchers at KI were published in journals that do not belong to the ISI Subject Category ‘oncology’. When looking at impact factors, only one of the 25 journals with the highest impact factor (2002) in which cancer researchers at KI published their results was categorised as oncology.

Table 3 - The twenty most common journals for cancer researchers at Karolinska Institutet

	Journal	Journal category
1	Journal of biological chemistry	Biochemistry & Molecular biology
2	International journal of cancer	**Oncology**
3	Bone marrow transplantation	Haematology Immunology **Oncology** Transplantation
4	PNAS	Multidisciplinary sciences
5	FEBS Letters	Biochemistry & Molecular biology Biophysics Cell biology
6	Blood	Haematology
7	Carcinogenesis	**Oncology**
8	Biochemical and biophysical research communications	Biochemistry & Molecular biology Biophysics
9	Cancer research	**Oncology**
10	Acta oncologica	**Oncology**
11	British journal of cancer	**Oncology**
12	European journal of cancer	**Oncology**
13	British journal of haematology	Haematology
14	Journal of immunology	Immunology
15	Oncogene	Biochemistry & Molecular biology Cell biology Genetics & Heredity **Oncology**
16	European journal of biochemistry	Biochemistry & Molecular biology
17	Anticancer research	**Oncology**
18	Genes chromosomes & cancer	Genetics & Heredity **Oncology**
19	Genomics	Genetics & Heredity
20	Experimental cell research	Cell biology

The 34 hearing researchers published their results in at least 53 different journals between 1994 and 2003. Five of the 17 most frequently used journals did not belong to the ISI Subject Category ‘otorhinolaryngology’. Further, eleven of the 17 most frequently used journals were not specifically devoted to hearing research. In total, 40% of the articles published by hearing researchers at KI were published in journals that do not belong to the ISI Subject Category ‘Otorhinolaryngology’ and 64% of the articles were published in journals not specifically devoted to hearing research. When looking at impact factors, none of the 25 journals with the highest impact factors in which hearing researchers at KI published their results was categorised as ‘otorhinolaryngology.

5.1.3 The choice of search strategy for article identification has an impact on evaluation and policy analysis of a research area

The different search strategies result in varying assessments with regards to the development of citations received (-13% – +27% *cancer* / -40% – +74% *hearing*), average impact factors (-4% – +8% / +43% – +126%) and total number of publications (+52% – +74% / +64% – +225%) (example in Figure 9). The seemingly most similar development indicated by the three strategies was seen for the total number of publications within the cancer research (Figure 10).

Figure 9 - Development of the average number of citations received the first two years post publication, for cancer publications at Karolinska Institutet between 1994 and 2001, using three different article identification strategies – MeSH-terms, journals and authors (indexed with the number for 1994 = 100).

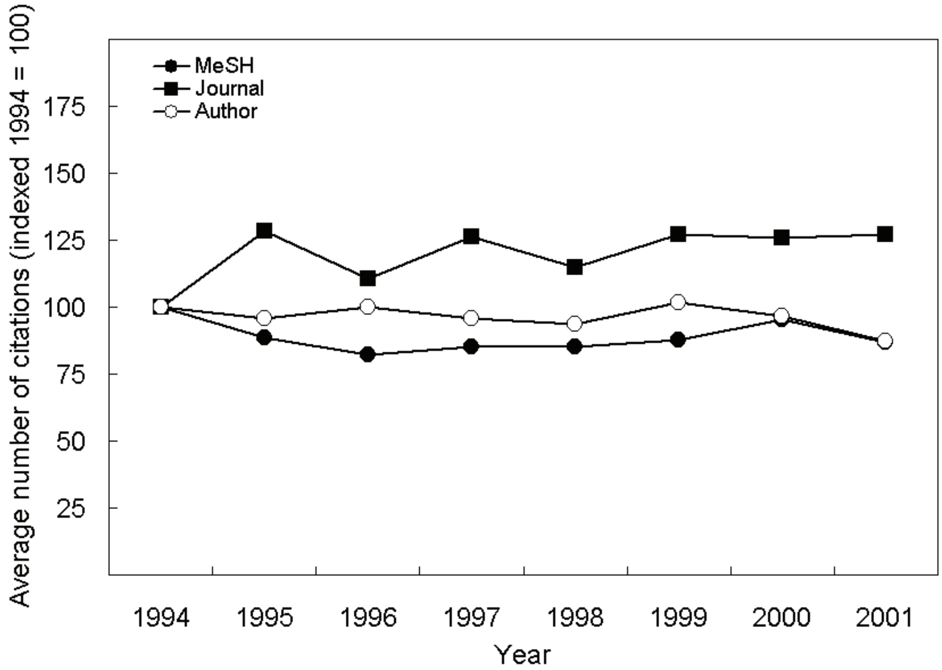
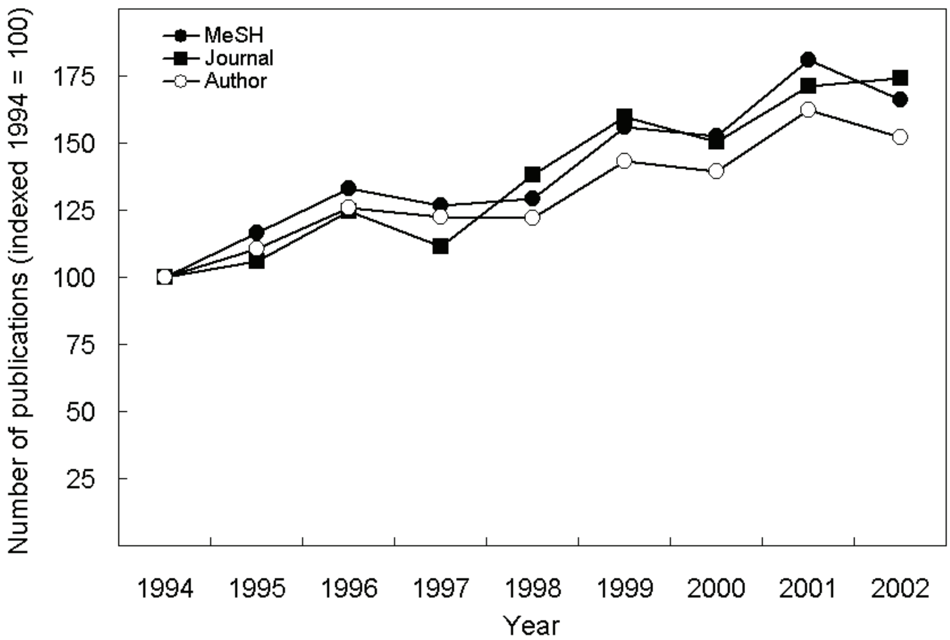


Figure 10 - Development of number of cancer research publications per year at Karolinska Institutet between 1994 and 2002, using three different article identification strategies – MeSH-terms, journals and authors (indexed with the number for 1994 = 100).



To summarize, the results of the first study of this thesis show that researchers often publish in journals which are not categorized as belonging to their own research areas. Furthermore, the choice of search strategy for article identification has an impact on evaluation and policy analysis of research areas.

5.2 Study II

The aim of the second study of this thesis was to assess how well university-industry collaborations can be identified and described using co-authorship data.[122]

5.2.1 Both co-authorship and funding indicators provide incomplete results in assessments of industry-university collaboration

In study *II* co-authorship data and industrial funding to a medical university were compared. In total 436 companies were identified through the two methods (Table 4). The results show that one third of the companies that provided funding to the university had not co-authored any publications with the university. Moreover, the funding indicator only identified 16% of the companies that had co-authored publications with the university.

Table 4 - Companies identified through co-authorship and/or funding

	Funding (+)	Funding (−)	*Total*
Co-authorship (+)	64	341	*405*
Co-authorship (−)	31	0	*31*
Total	*95*	*341*	*436*

Most of the 31 companies that had provided funding but not co-authored any publications were small- or medium sized biotechnology companies, but also larger pharmaceutical- and tobacco companies were among these companies. The 341 companies co-authoring publications without providing funding included a wide variety of companies. The ten companies that had co-authored more than ten publications without providing funding primarily were medical technology- and pharmaceutical companies.

5.2.2 Co-authorship and funding indicators does not always indicate the similar development of collaborations

A case of conflicting trends between funding and co-authorship indicators was observed in the study as shown in Figure 11 and Figure 12. The number of co-authored publications during the years 1999 and 2002 decreased (Figure 11).The corresponding decrease was not found in the level of funding. These conflicting trends became even more evident when looking at the average number of co-authored publications per million SEK (adjusted for inflation) in research funding from 'PharmaInc' to KI (Figure 12). Even though there has been a general increase in funding per publication it does not explain the rapid change in average funding per publication.

Figure 11 - Development of funding from and co-authored publications with 'PharmaInc'

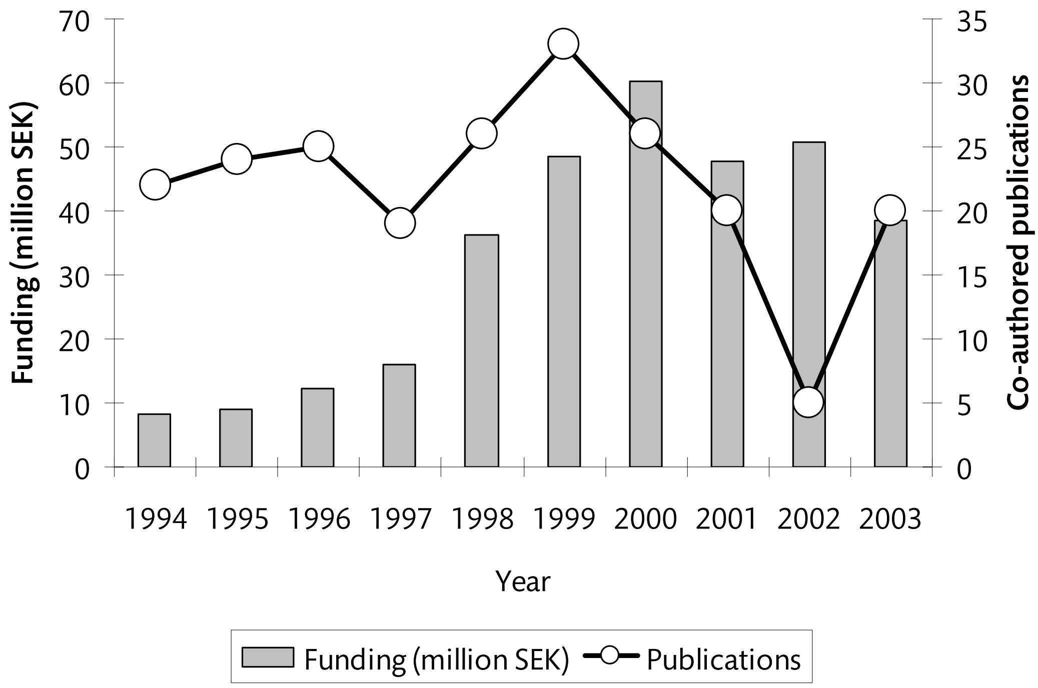
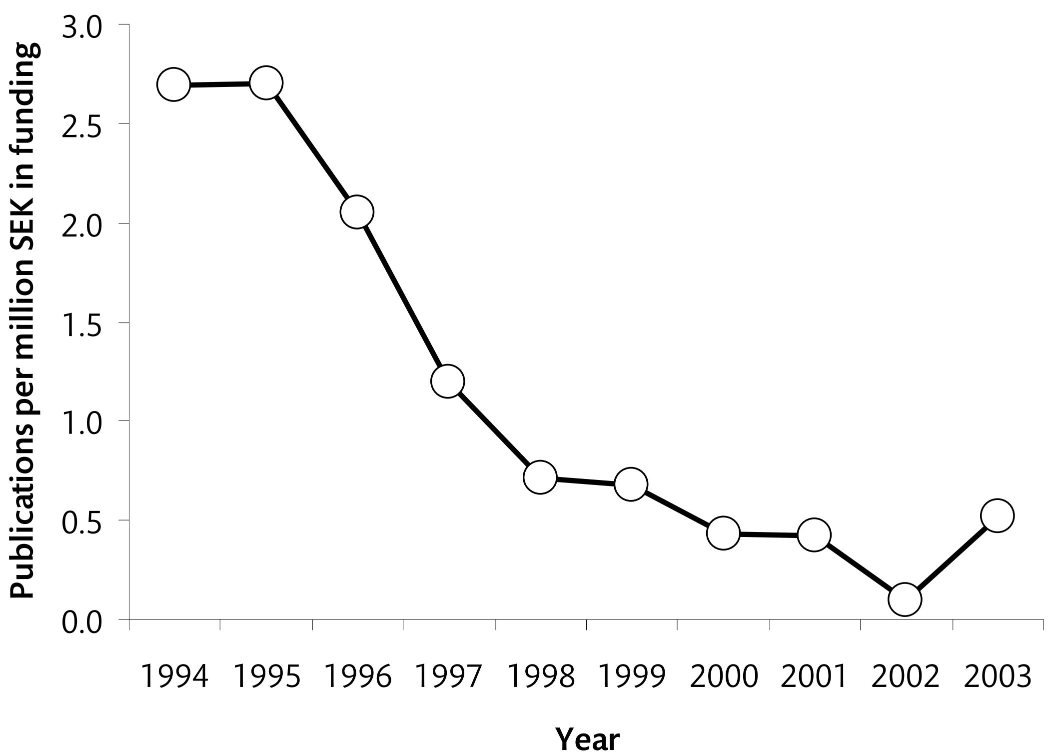


Figure 12 - Average number of co-authored publications per million SEK (adjusted for inflation) in research funding from 'PharmaInc' to Karolinska Institutet



5.2.3 The distribution of the number of companies by number of co-authored publications is skewed

Seventy-two percent of the companies identified through co-authorships had only co-authored one or two publications with the university during the years 1995-2003. Sixteen percent of the companies had co-authored more than five publications, nine percent more than ten publications and one percent (six companies) had co-authored more than fifty publications during the same time period.

In summary, the results of study *II* demonstrate that both co-authorship and funding are incomplete indicators of collaboration. In some cases the two indicators might even signal contradictory trends.

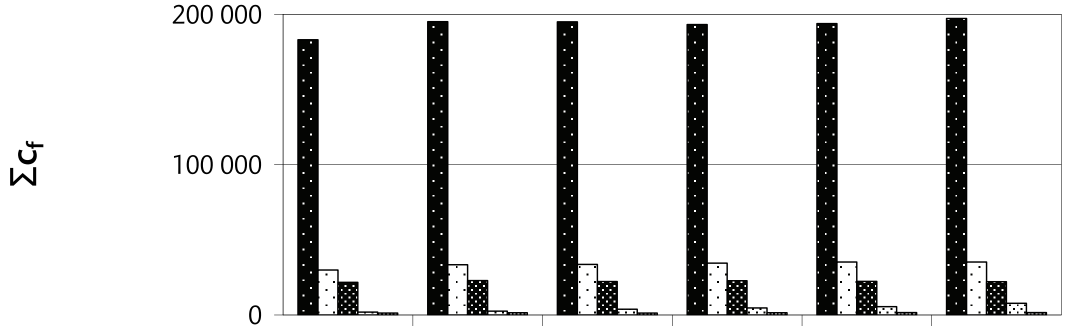
5.3 Study III

The aim of study III was to introduce new bibliometric indicators that deal with flaws related to the current state-of-the-art indicator.[123]

5.3.1 Assigning equal weight to all publications

When calculating the state-of-the-art indicator, citation rates are not normalized on the level of individual publications, but on a higher aggregation level where the average citation rate of a researcher, group or department is compared to the average citation rate of the fields in which the researcher or group has published. This way of calculating gives more weight to older publications (particularly reviews), published in fields with dense citation traffic. In order to give each publication equal weight the normalization should take place on the level of the individual publication. The calculation of such an *item oriented field normalized citation score average* - $\bar{c}_f$ - is described in detail in the third study of this thesis.[123] Here, instead of first calculating the actual average citation rate and then divide that with the average expected citation rate, each publication (or item) is normalized individually (hence 'item oriented'). The field normalized citation scores can also be summed in order to calculate a total field normalized citation score (Σc_f) for a research group, university or country. An example of Σc_f calculation for publications from five countries 1998-2003 is shown in Figure 13.

Figure 13 - The total field normalized citation score 1998-2003 for 1 416 912 publications in 57 life science related subject categories, co-authored by researchers affiliated with organisations in USA, Germany, France, China and Russia. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.



	1998	1999	2000	2001	2002	2003
USA	183 208	195 151	194 999	193 163	193 781	197 374
Germany	29 889	33 385	33 630	34 453	35 268	35 202
France	21 670	22 829	22 281	22 765	22 452	22 149
Peoples R China	1 942	2 602	3 827	4 755	5 580	7 704
Russia	1 368	1 426	1 343	1 460	1 526	1 577

Year

5.3.2 Controlling for skewed and differing distribution of citations over publications – citation *z*-score

The distribution of citations over publications differs between 'normalization groups' (publications of a specific type, within a specific research field, published a specific year). Therefore not only does the average citation rate differ, but also the standard deviation. It could thus be argued that it would be appropriate to use a *z*-score in the normalization procedure. A *z*-score expresses how far a value is from the population mean in terms of the number of standard deviations by which it differs.[124] A second issue that needs to be dealt with is that the distribution of citations over publications is highly skewed. The skewed distribution has been seen for fields and journals, as well as for individual scientists.[68] Since the distribution is skewed it could be an alternative to use the geometric mean or median value as comparison when calculating normalized citation rates. Another alternative is to make normalizations using logarithmically transformed citation rates. We chose not to use the median value since the median for most normalization groups would be just one or a few citations. The step between for example zero and one or one and two would have a very large impact on the normalized citation rates, even though the difference between the groups would not be significant. For this reason the median value was not used in the study. The geometric mean was considered but abandoned in favour of using logarithmically transformed citation rates. This in order to reduce the impact of the very skewed distribution not only on the average for the normalization group, but also on the average normalized citation rate for a group, department or university.

Combining the observations above, it could be argued that it would be appropriate to use a logarithm based *citation z-score* as a complementary indicator to the item oriented field normalized citation score average. A citation *z*-score would compare the logarithm of the number of citations which a publication has received with the mean and standard deviation for the logarithms of the citation rates for all the corresponding reference publications. Details on the calculation are given in study *III*.[123]

In comparison to the crown indicator, which distribution starts to approach normal on aggregated levels (such as distribution of research groups over classes of crown indicator values), the distribution of field normalized logarithm-based citation *z*-scores ($c_{fz[\ln]}$) over publications starts to approach normal distribution already *within* low aggregation levels such as research groups (Figure 14). $c_{fz[\ln]}$ values are approximately normally distributed *within* department or university level (Figure 15). On aggregated levels the citation *z*-score is approximately normally distributed on the level of research groups (Figure 16).

Figure 14 - Distribution of publications over classes of $c_{fz[\ln]}$. The figure is based on 453 publications by the 20 research groups at a department at Karolinska Institutet, published between 1998 and 2003. Each line represents the publications by a research group. The lines are placed on top of each other. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.

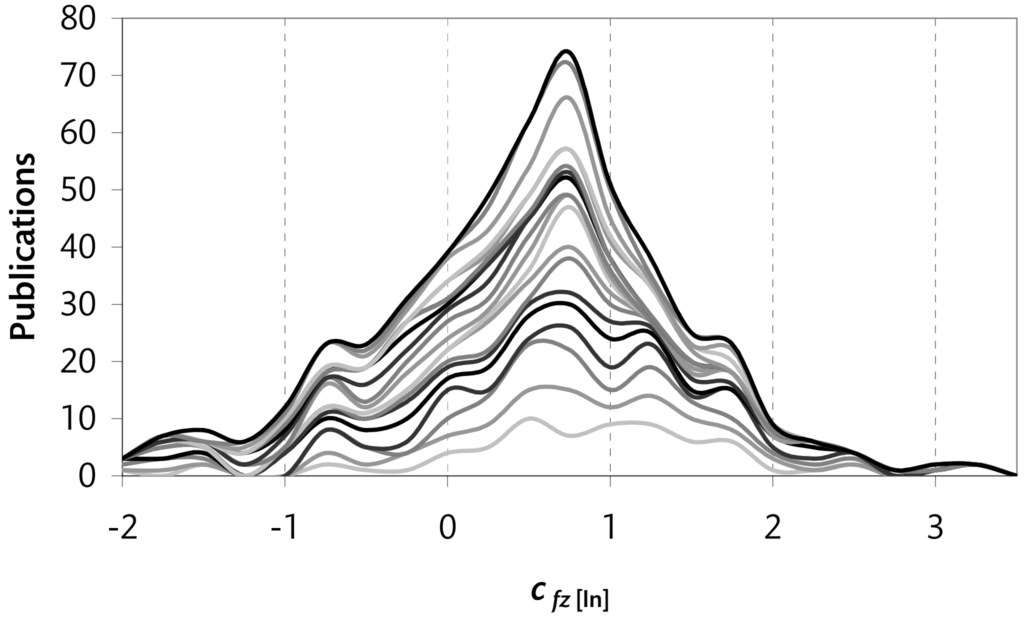


Figure 15 - Relative distribution of publications over classes of $c_{fz[\ln]}$ for a department at KI, KI as a whole and all Swedish publications within 57 research areas (1998-2003)

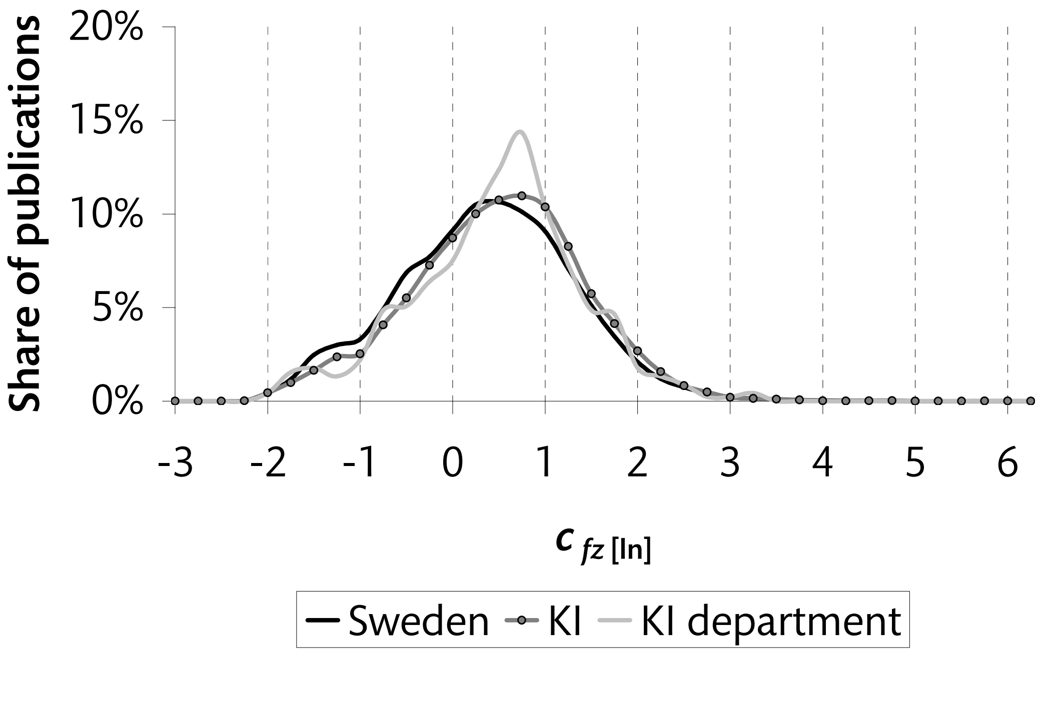
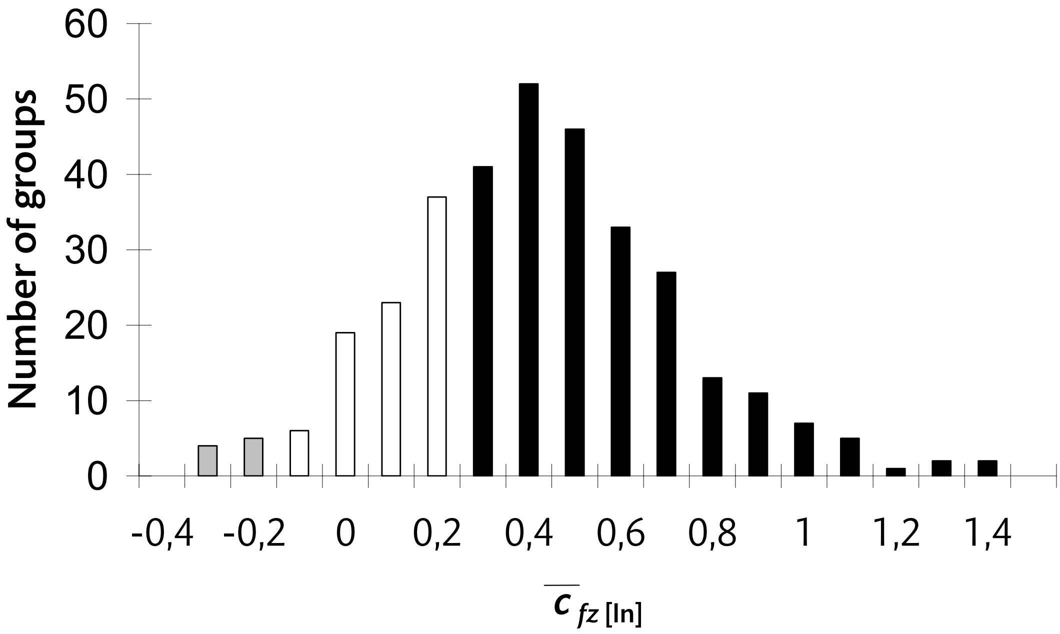


Figure 16 - Distribution of research units over classes of citation *z*-scores. The figure is based on data for 334 research units at KI (at least 30 publications 1995-2004). Bars including groups with citation *z*-score < -0.2 are coloured grey, citation *z*-score -0.2 and +0.2 white and citation *z*-score > +0.2 black. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.



To sum up, the results of study III show that the current state-of-the-art bibliometric indicator has flaws and some of these could be dealt with using a citation *z*-score.

5.4 Study IV

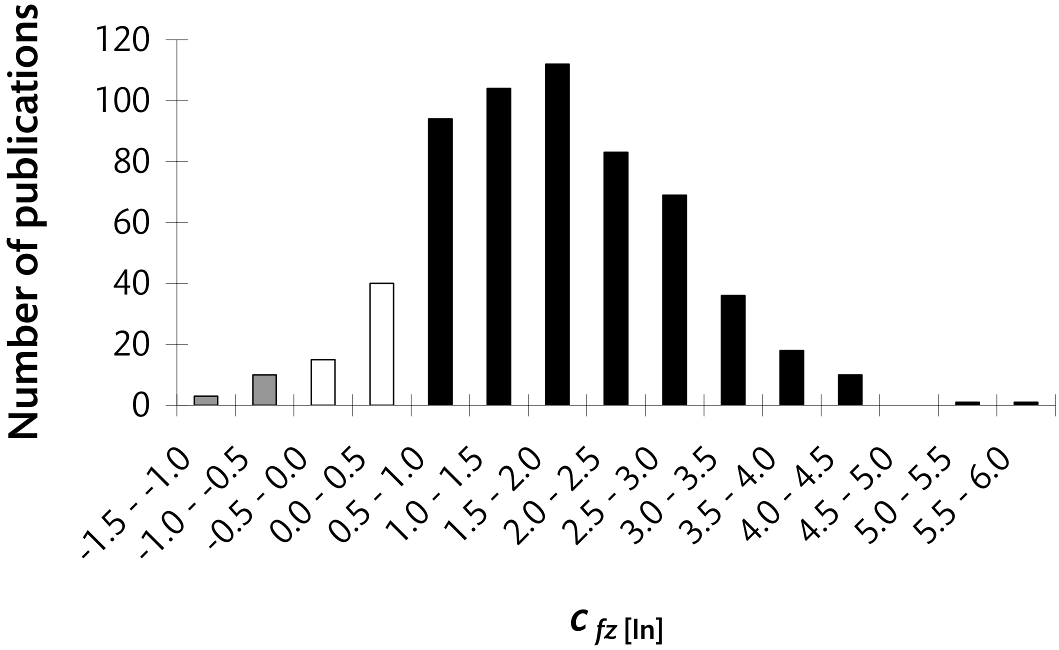
The objective of the last study of this thesis was to assess whether publications of importance for improving the health system and its technologies are highly cited intra-scientifically.[125]

5.4.1 Publications used as sources in a Swedish system for identification and early assessment of new methods in health care are also highly cited within the scientific community

The results of the study showed that the publications used as sources in SBU Alert were highly cited compared to publications of the same type, from the same year, published in journals within the same field. On average they are cited eight times more than the world average ($\overline{c}_f = 7.79$). The citation *z*-score is 1.7 standard deviations above the world average ($\overline{c}_{fz[\ln]} = 1.69$). The high bibliometric score is consistent over the whole assessed time period, with the highest value seen for publications used as sources in SBU Alerts published in 2003 ($\overline{c}_f = 10.3; \overline{c}_{fz[\ln]} = 1.9$) and the lowest for SBU Alerts published in 2002 ($\overline{c}_f = 4.7; \overline{c}_{fz[\ln]} = 1.4$). Only 12% of all articles had received fewer citations than the world average within their field.

The sensitivity (or recall) of using positive normalized citation score to identify articles used as sources in SBU Alert is 0.88. 73% of the publications had a citation *z*-score more than one standard deviation above the world average. The distribution of publications over classes of $c_{fz[\ln]}$ is shown in Figure 17. Only two of the 50 SBU Alerts used publications as sources that on average have received fewer citations than the world average. No SBU Alert used sources with citation z-score below the world average.

Figure 17 - Distribution of publications over classes of $c_{fz[ln]}$



5.4.2 The science base in SBU Alert is dominated by the United States and the United Kingdom

Co-authors of the publications cited in SBU Alerts 2001-2004 most frequently had an affiliation in the United States (31%) or the United Kingdom (12%). The figure for USA was in line with the expected (0.9), while the UK was overrepresented by 1.6 times. The highest overrepresentation, according to the method used, was seen for Sweden (5.1 times more than expected), Denmark (2.4) and the Netherlands (2.3). Japan (0.2) and Taiwan (0.2) were the two most underrepresented countries out of countries with at least one publication used as source in SBU Alerts. It could be noted that researchers affiliated with organisations in India and Russia had not been co-authors of any publications used as sources.

In summary, the results of the fourth study of this thesis showed that publications used as sources in a Swedish system for identification and early assessment of new methods in health care are also highly cited within the scientific community. It was also noted that important results are published in high impact medical journals as well as in field-specific journals with low impact factors.

5.4.3 Articles used as the scientific basis for SBU Alerts are published in high impact medical journals as well as in field-specific journals with low impact factors

The New England Journal of Medicine was the most frequently used source for information, followed by Circulation and the Lancet (Table 5). There was a considerable range in the impact factors of the journals.

Table 5 - Journals in CI, being cited at least five times in SBU Alert reports 2001-2004. I_{ISI}: ISI journal impact factor 2005.

n	Journal	I_{ISI}
46	New England Journal of Medicine	44.0
35	Circulation	11.6
30	The Lancet	23.4
17	Radiology	5.4
16	JAMA	23.3
12	International Journal of Oral & Maxillofacial Implants	1.4
11	Clinical Oral Implants Research	1.9
9	Pediatrics	4.3
8	American Journal of Cardiology	3.1
8	British Journal of Surgery	3.7
8	Clinical Therapeutics	3.0
8	European Journal of Vascular and Endovascular Surgery	2.0
7	Diseases of the Colon & Rectum	2.3
7	Journal of Vascular Surgery	3.2
6	Annals of Thoracic Surgery	2.2
6	Archives of Internal Medicine	8.0
6	British Medical Journal	9.1
6	European Heart Journal	7.3
6	Movement Disorders	2.8
5	Artificial Organs	1.9
5	Brain	7.5
5	European Journal of Heart Failure	3.5
5	European Urology	3.5
5	Gastroenterology	12.4
5	International Journal of Artificial Organs	0.9
5	Journal of Clinical Endocrinology and Metabolism	6.0
5	Journal of Thoracic and Cardiovascular Surgery	3.7
5	Neurology	4.9

6 Discussion

The results of this thesis show that uncritical assessments of research areas based on rudimentary article identification strategies or collaboration analyses based solely on co-authorship data may lead to misinterpretation of the development and thus provide incorrect information for decision-making. At the same time, a correct use of refined bibliometric indicators may provide valuable background information for decision makers.

6.1 Is it better or just the same?

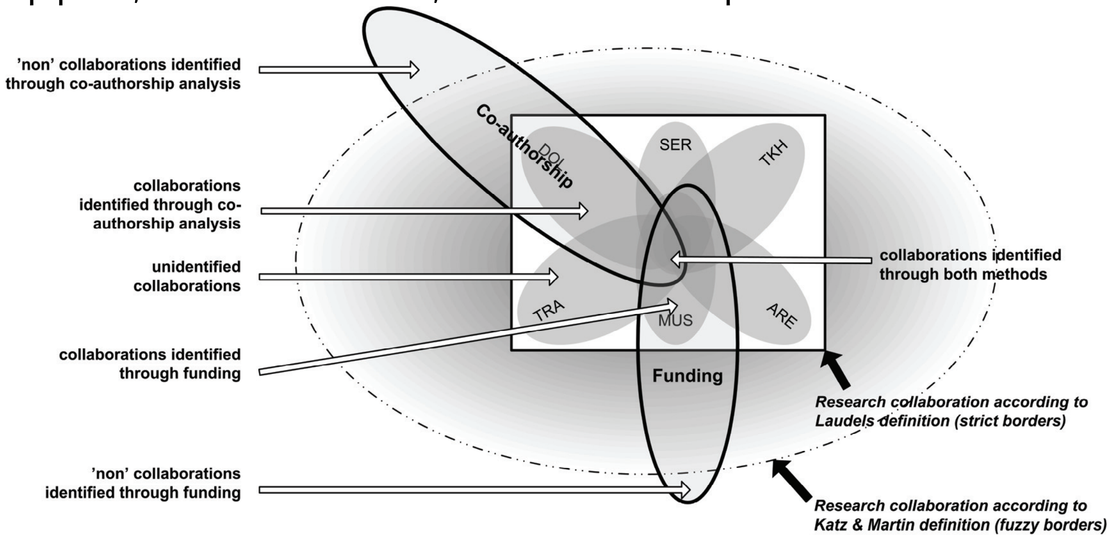
Different search strategies result in varying assessments with regards to the development of citations received, average impact factors and total number of publications (*I*). Therefore the choice of search strategy for article identification can have an impact on evaluation and policy analysis of research areas. As a consequence, uncritical analysis based on rudimentary article identification strategies may lead to misinterpretation of the development of research areas and thus provide incorrect information for decision-making.

Using a journal based search strategy limits the number of identified articles and increases the possibility that relevant articles are not included in further analysis (*I*). These findings are in line with the results of a study by Ugolini and co-workers.[100] They showed that using a journal based search strategy to identify cancer publications has several shortcomings. Many cancer research articles are published in non-oncology journals and some policy related analyses such as country rankings of mean impact factors and publications per GDP show different ranking orders when using a broader search strategy.[100] In a case study of ophthalmology research by van Leeuwen, a partial overlap between journal categories and field-specific addresses was shown.[28] He showed that the journal strategy shared about 70% of the references with the field-specific address strategy. These findings support the results from our study where the journal category does not cover all important articles within the areas. These results support the conclusion by Lewison that a search strategy limited to journal categories, risks omitting many relevant articles.[99] At the same time, less specific search strategies increase the likelihood of irrelevant articles being included. In finding the best identification strategy one must come to a decision based on the relative importance of recall and precision in each particular case.[126] Since the MeSH-strategy in general had high recall as well as precision, this strategy is to be recommended over the other two. At least the identification of research areas should be performed on the level of individual publications rather than on the level of the journal. The author based strategy can be useful for assessing or mapping developing areas or areas with unclear borders. In this case the definition of a complex area such as cancer would be “cancer research is what cancer researcher does”. This definition may be useful for mapping or internal assessment purposes, but becomes difficult when comparing cancer research between two universities, or trying to normalize citation rates within the area. A fixed definition would probably be more useful.

6.2 Collaboration uncovered

Co-authorship and funding are incomplete indicators of collaboration (*II*). One third of the companies that provided funding to the university had not co-authored any publications with the university. Further, the funding indicator only identified 16% of the companies that had co-authored publications. This can be explained by the idea that collaboration should be considered as a complex phenomenon with fuzzy borders.[70] Therefore input- (like funding) and output indicators (like co-authorships) will only reveal parts of the actual collaborations. According to Laudel, primarily collaboration involving a division of labour result in co-authored publications and people are included as collaborators only if they conduct research activities.[69] Only to provide funding is not included in Laudel's definition of research collaboration, nor in any of the six specific types of research collaboration. This is reasonable since we, for example, generally do not consider funding agencies as collaborators in research projects. Does this mean that industrial funding to universities should be considered as a new type of research collaboration? In the second study of this thesis it was suggested that funding should not be considered as a type of research collaboration in itself, but rather as an additional indicator of research collaboration.[127] When a company is involved in setting the research agenda for a research group (could be considered as collaboration of the type 'mutual stimulation'), provides access to research equipment or is involved in the transmission of know-how, this collaboration is not likely to result in co-authored publications.[69] Though these types of collaboration probably involve that the company provides funding to the project. Thus funding could be an indicator of collaboration which is often not detectable through co-authorship analysis. Still it is important to note, that since funding is not collaboration in itself, it is possible that a funding indicator will show collaborations that are not 'real'. A summary of how co-authorship and funding could measure different aspects of research collaboration is shown in Figure 18.[127]

Figure 18 - Summary of how co-authorship and funding could measure different aspects of university-industry collaboration. DOL: Collaboration involving a division of labour; SER: Service collaboration; TKH: Transmission of know-how; ARE: Provision of access to research equipment; MUS: Mutual stimulation; TRA: Trusted assessorship.



6.3 Lifting the crown

The current state-of-the-art bibliometric indicator has flaws, but some of these could be dealt with using a citation *z*-score (*III*). The three new indicators suggested in this thesis build on earlier research which has shown how citation rates could be normalized for research field, publication year and document type.[65-67] The *item oriented field normalized citation score average* ($\bar{c}_f$) is an incremental improvement of the current state-of-the-art indicator (the crown indicator). It differs from the crown indicator in so much as it assigns equal weight to each included publication. That normalization takes place on the level of individual item also makes it possible to calculate total field normalized citation score (Σc_f). The more radical improvement (or complement) is the *item oriented field normalized logarithm-based citation z-score average* ($\bar{c}_{fz[\ln]}$ or *citation z-score*). This indicator assigns equal weight to each included publication and takes into account the citation rate variability of different fields as well as the skewed distribution of citations over publications. In comparison to the crown indicator, which distribution starts to approach normal on aggregated levels (such as distribution of research groups over classes of crown indicator values), the distribution of field normalized logarithm-based citation *z*-scores ($c_{fz[\ln]}$) over publications starts to approach normal distribution already *within* low aggregation levels such as research groups (Figure 14, p. 35). Since the citation *z*-score is based on logarithmic transformations of citation rates, extreme values have less impact on this indicator. This is both the strength and the weakness of this indicator since one often is interested in precisely those extreme values when assessing research.[128] The citation *z*-score thus provides a complementary view to other indicators which are more heavily influenced by extreme citation rates (e.g. $\bar{c}_f$) or solely concern very highly cited publications (e.g. share of publications in top 1% of the citation distribution). Therefore it seems appropriate to provide information on both types of indicators as input to informed peer review.

6.4 Breakthroughs and bibliometrics

The publications that are used as sources in SBU Alert are cited eight times more often than the world average (*IV*). This is a very high value considering that in assessments of research groups of similar size (~50 publications per year) an average citation rate of three times the world average signals excellence.[48]

There was no a priori theory on whether publications used as sources in a system for identification and early assessment of new methods in health care also should be highly cited. Intuitively this is to be expected since high quality publications often are highly cited and it is reasonable to believe that publications serving as basis for technologies judged as promising and important are also of high quality. It has been argued that there is bias against more clinically oriented research in bibliometric assessments based on the CI.[56, 112, 113] The bias would exist since the CI are based on citations given in scientific journals and not on documents more closely related to patient care, such as guidelines or systems for identification of emerging technologies in health care. The results of this thesis contribute to this discussion. It is possible that the use of field normalized bibliometric indicators (*IV*) reduces some of the bias against clinical research earlier shown to exist when using for example impact factors.

Articles in a few high impact medical journals often served as the scientific basis for SBU Alert (*IV*). However, several of the articles were published in field-specific journals with low impact factors (Table 5, p. 38). This finding argues against mechanical use of impact factors in assessment of clinical research. Impact factors may hide the very variations they are believed to display and are also much dependent on the composition of document types in a journal and the subject category it covers.[56]

Most articles used in SBU Alerts were published by authors based in the United States or the United Kingdom (*IV*). Publications from Sweden and from countries with similar health systems, such as Denmark and the Netherlands, were overrepresented as science base in SBU Alert. This is natural since local adaptation often is necessary when turning research into practice.[129] It is also in line with a study on the scientific basis of British clinical guidelines which showed UK papers to be represented two and a half time more than expected.[130]

Eight percent of the sources (published 1995-2004) used in SBU Alert were articles published in journals not covered by the CI (*IV*). It is not unreasonable to believe that some medical researchers may publish considerable parts of their work in such journals and thus be unsuitable for bibliometric assessments based on CI. The results also show that the journals outside the scope of CI play an important role in medical research.

6.5 Methodological considerations

This thesis may, in itself, be considered as a discussion of methodological consideration related to bibliometrics.

In all four studies substantial work has been put into data correction and validation. Authors who have changed names have been identified, as have alternative spellings (and misspellings) of their names (*I*). Organization names have been corrected to the level of 'current parent company' (*II*) and company publications have been identified through manual assessment of 17 117 addresses (*II*). For data on research group level, data was used from a manual validation of publications by over 600 research group leaders (*III*). This validation process was conducted during six months in 2005. A manual matching of all 953 references given in SBU Alerts 2001-2004 with items in CI was also conducted within the scope of this thesis project (*IV*).

In the first study the use of expert opinion as 'golden standard' is a weakness. In this case two cancer researchers assisted in the assessment of whether publications should be classified as oncology. It is likely that other researchers would have made different categorizations. There is room for improvement of the list of MeSH-terms used for article identification. The list in the study was developed to include relevant and specific terms (high recall and high precision). Some terms were deemed as too general and therefore excluded. Additional work on the list as well as a different group of researchers selected as 'golden standard' might have resulted in altered assessments of the recall and precision of the article identification strategies. However, it is not likely that it would have changed the results, since the MeSH-based strategy had a much higher recall than the journal based strategy.

Another limitation of the first study is that non-normalized indicators were used. It is not certain that the varying article identification strategies would have the same impact on assessments using modern indicators such as the crown indicator or citation *z*-score. However, the three methods used in the study (total number of publications, average impact factor and citations two years post publication) are in many cases standard indicators in practice. The study is, therefore, raising additional doubt on the usefulness of these bibliometric indicators, especially of the two latter.

Some limitations of second study involve the use of funding as comparative indicator. Even though this indicator proved to be useful it would have been valuable to add a qualitative method to the study to triangulate the phenomenon under study. This could have provided interesting details on how and why collaboration which does not involve co-author or funding takes place. An alternative 'golden standard' could also have assisted in assessing the recall and precision of co-authorship as a measure of collaboration.

The three suggested indicators have been used on example data (*III*). Their usefulness has primarily been shown from a theoretical point of view. Publications related to promising health care technologies in general received high indicator values (*IV*). The validity of the indicators needs to be further addressed. First of all the correlation between peer judgments and the indicators must be investigated. Second of all their predictive value must be attended to. It would have been valuable if this had been done already within this thesis project. Unfortunately it was impossible for practical reasons, within the scope of this project. Nevertheless, an important first contribution is simply to develop and explain the new indicators.

A main drawback of the fourth study was that all cited items were considered as equal regardless of whether they were referred to in a negotional or confirmatory way. Additionally it would have been an advantage to conduct additional analysis on publications similar to SBU Alert in organizations from other countries. Study *I* and *II* would also have benefited from international comparisons. With regard to the last study another important methodological consideration is that the conclusions are drawn based on the assumption that technologies selected for SBU Alert are to be considered as 'deemed as promising by clinical experts'. In addition, the predictive value of 'deemed as promising by clinical experts' for future patient benefit has not been assessed. If the predictive value is low, this would decrease the merit of the study.

6.6 Reflections and future research

6.6.1 Use and misuse of bibliometric indicators

Bibliometric methods can be used to test a vast number of hypotheses of scientific and practical importance. From a scientific point of view the methods can assist in the description and exploration of both the social and cognitive development of science. Their main applications are, though, probably as tools for science policy and research management. From a science policy perspective it is valuable to monitor the development of supported science. The results of interventions such as increased or reduced investments, or new research strategies could be observed in terms of research volume (number of publications), international influence (Σc_f) average influence (citation z-score) or collaboration (co-authorship). There is also a considerable range of application areas of bibliometrics for research managers. An almost automatically updated view on the status of the research that is carried out within the organisation is a first step. This can be supplemented by, for example, more detailed views on the strengths and weaknesses of research areas or the effect of management decisions on research practice.

As stated earlier the results of bibliometric analyses should be viewed with a critical mind and eyes open to indicator limitations and alternative explanations to measurements. Each bibliometric indicator has a specific limitation. This is one of the reasons why these types of evaluations should incorporate more than one indicator.[24] There is also substantial room for manipulation by selection, weighting and aggregating indicators.[49] It is possible that the average citation rate of a research unit is high because one article of the group is highly cited, with other publications receiving very few. Concerns relating to this skewed distribution are most critical if the number of publications is small, i.e. less than 50 publications. Misuse of bibliometric indicators includes neglecting their limitations. Grupp and Mogee exemplified four such misuses: (1) Reliance on single indicators; (2) The use of an indicator that is inappropriate for the studied technology, system, or stage of the R&D process; (3) Drawing conclusions which are too strong, given the 'indicative' nature of indicators; and (4) Making inferences that are inappropriate, based on the indicator and its relationship to the phenomenon of interest.[49] van Leeuwen warned that even though the creator and direct user of a bibliometric study may be confident in what conclusion that could be drawn based on the results, it is possible that other users who are not knowledgeable in bibliometrics in general or in the specific study may overlook the limitations of the study.[131]

Direct linking between bibliometric assessments and resource allocation can be dangerous. Any system used to assess research that affects money and/or prestige is likely to affect the behaviour of researchers. When bibliometric indicators are related to resource allocation or other types of sanctions, they become 'reactive measures', i.e. people will react to the implementation of the measures by altering their behaviour.[38] On the surface this is of course what the decision makers want (to have an impact on the researchers), but it can result in 'goal displacement' i.e. high scores in the measures become the goal rather than a means of measuring whether an objective has been achieved.[132] Another unintended consequence might be that the research process itself is modified as researchers adapt their behaviour in response to the evaluation. It has for example been noted that based on the focus on bibliometric evaluations, scientists have changed their publication strategy by publishing more in international journals which has made it more difficult for nationally oriented journals to attract enough manuscripts. Although a stronger international orientation in publishing can be desirable, it could lead to the neglect of nationally important topics. It could also be the case that the utility of bibliometrics as a research assessment tool is diminished by incorporating it to much into resource allocation systems. The theoretical foundation of the validity of citation based performance indicators is that in some sense "good research get highly cited". If scientist knows that their (and their colleagues) are being assessed by citation rates and that money is being allocated, this is likely to promote behaviours such as *"if you cite me, I'll cite you"* or "*you'd better cite me, or else…*" These behaviours will increase the citation rates of some people without increasing the importance or influence (or even less so the quality) of their work.

The area of 'measurement burden' should also be discussed. It must be remembered by managers and policy makers that "*if the aim of measurement is to lead to an improvement of some process one must be certain that the cost of taking that measurement does not outweigh the gains to be made through knowing it.*"[46] This in general favours indicators that are part of the everyday running of an organization. Bibliometric indicators, being based on an important everyday output of science can be considered as such indicators, but the collection, correction and validation of data must be performed in such a way that it limits the amount of double work that must be performed by individual scientists when reporting their progress to different actors. This also calls for an increased collaboration between bibliometric researcher and openness towards sharing corrected data.

6.6.2 Should citations be interpreted differently in medicine than in the social sciences?

As was mentioned in the introduction there is an ongoing debate in the bibliometric community on whether researchers (1) give references to register intellectual property and peer recognition and that counting citations thus measures intellectual influence *or* (2) that references are primarily used for rhetorical purposes, i.e. to advance the interests of a researcher, defend his or hers claims against attack, convince others and thus gain a dominant position in their scientific community.[39-41] This dichotomy might be the result of differing scientific traditions between fields. It could be hypothesized that researchers in more experimental and quantitative fields, such as the natural sciences or medicine, are prone towards referencing papers that were first with reporting a new method or uses of a certain dataset. Thus the reference is given based on priority or data quality. In the social sciences and more qualitative fields references are more often given to strengthen the argument of the author. The references do not have to be made to a certain methodology or result, but rather to the argumentation or model presented by a certain researcher. If this is true, it would increase the interpretability of citation based indicators, but also reduce the applicability of combining results from different fields into single indicator values (e.g. for a university). This does not reduce the importance of field normalization since the normalization procedure enables comparisons of related disciplines over time. In the actual assessment of individual papers or researchers it is worth to bare in mind the words of Sidney Brenner "*what matters absolutely is the scientific content of a paper and that nothing will substitute for either knowing it or reading it*".[133]

6.6.3 Does citation based performance indicators measure 'quality'?

Citation based indicators are not likely to measure one concept such as 'quality' and a wider spectrum of bibliometric indicators even less so. Instead bibliometric indicators may be valid indicators of different concepts under specific conditions. For example, consider two of the indicators proposed in this thesis, $\overline{c}_f$ and Σc_f. The $\overline{c}_f$ is in this case likely to be related with the concept 'quality' or other concepts which are not heavily influenced by sheer size. The Σc_f on the other hand is more likely to be related to concepts such as 'influence'. Similarly the citation *z*-score, which reduces the influence of single extremely highly cited publications, is more likely to be related to the 'quality' concept. People, groups or universities that are thought of as 'excellent' will probably be associated with extremely highly cited publications.

6.6.4 Is there only room for excellent science?

In one sense, single publications – or rather the discoveries presented in those publications – may have a larger impact on the scientific community or society as a whole, than the rest of the publications from an organization or a country during the same year. Still, it is important to note that these single paradigm altering publications are based on the work of others. Without the global production of 'normal science' (and the history of science) there would hardly be any major discoveries at all. I therefore support the idea of science as a collective effort and the old notion of "standing on the shoulder of giants". Those who want to take bibliometrics hostage in order to put all resources into only one or a few groups believing that they are 'best' should remember that as soon as the 'worst groups' are removed, someone else will be worst. Continuing this process only the 'best' will be doing research and – sorry to say – becoming the 'worst'.

6.6.5 When will bibliometrics disappear?

Bibliometrics as we know it today is in many ways threatened. For example, will a widespread misuse of bibliometric methods diminish the confidence put in it and reduce all resources spent on both development and analysis to zero. If the way we communicate science changes fundamentally, bibliometrics will also have to change or sink into oblivion. The main carriers of scientific communication today are the articles published in peer-reviewed journals, which might not last for long. One should remember that the original reason for starting to publish research findings in journals instead of books is suggested to be that "*there were too many books*".[6] When will there be *too many journals?* In the future journals will be replaced by something else. If this 'something else' will be more or less measurable than journals, -who knows? At the same time it is likely that all of this will take time. The journal article is built into so many structures that it is impossible to reduce its importance in haste. For example, articles are used for establishing priority, communicating science and building up the résumés of most distinguished scientists.

6.6.6 Impact beyond the impact factor?

The subtitle of this thesis was given for three different reasons. First of all to stimulate other scientists to read it. Second, the new indicators described in this thesis, as well as the validity work done in the other sections, show that bibliometric methods have developed beyond the impact factor. There are indicators available that can be used for more valid research assessments. Thirdly, I think it is time that we start to move beyond bibliometrics. The fourth study in this thesis was one step to bring bibliometrics closer towards impact of research beyond the scientific community. This was just a first step and more work is necessary in order to establish where the borders of bibliometrics are located. Beyond those borders we must find other indicators measuring the true impact of research.

7 Conclusions and policy implications

The utility of bibliometrics as a research assessment tool has been explored and developed through the results of this thesis. This has been done by identifying the limitations as well as new possibilities, and includes four conclusions:

- The choice of search strategy for article identification has an impact on evaluation and policy analysis of research areas.
- Both co-authorship and funding indicators provide incomplete results in assessments of industry-university collaboration. Uncritical use of the two indicators may lead to misinterpretation of the development of collaborations and thus provide incorrect information for decision-making.
- The current state-of-the-art bibliometric indicator has flaws, but some of these could be dealt with using a citation *z*-score. Still, it should not be used as a sole indicator of research performance but rather as one of many indicators used as input for informed peer review.
- Publications used as sources in a Swedish system for identification and early assessment of new methods in health care are also highly cited within the scientific community. This increases the appropriateness of using bibliometric indicators in evaluations of clinical research.

There are some policy implications of these results.

First earlier views on the development of research areas should be reconsidered. For example, thoughts on the comparative strength of research within certain areas in different universities might not be accurate. When requesting new analysis of research areas it is recommended that they are performed by using research area assignment on the level of individual publications rather than journals.

Second, there is currently a trend of using co-authorship data as evidence in collaboration assessments. This is for example done when assessing the integration of the European research area, or the level of industry-university collaboration. The results of this thesis have shown that important collaboration may not result in co-authored publications. Further co-authored publications may not be the results of actual collaborations. Therefore co-authorship data should be used with a critical mind and requests for additional indicators based on complementing data sources be made which makes triangulation possible.

Third, there is still no perfect citation based performance indicator available. The normalized indicators represent an improvement compared to basic bibliometric indicators, but there are still flaws associated with them. Policy makers should demand analysis using normalized citation rates and also instruct (and help) the bibliometric community to continuously work on the validity and reproducibility of these indicators.

Last, as was shown in the fourth study, bibliometric indicators may be utilised in research assessments, even in more applied (or clinical) areas.

In conclusion, bibliometric methods can be a useful tool in research assessment. The validity and appropriateness of using the methods may be increased by the use of refined bibliometric indicators. Uncritical assessments of research areas based on rudimentary article identification strategies or collaboration analyses based solely on co-authorship data may lead to misinterpretation of the development and thus provide incorrect information for decision-makers. Further research on the validity of bibliometric indicators and the implementation of such indicators in practice is needed in order to contribute to the development of science world wide.

8 Postlude

…it was a great view. This work on exploring and developing the utility of bibliometrics as a research assessment tool has taken its stance in the research performed by scientists before me. Standing on shoulders of giants I have tried to tell you what I saw from up there. The bibliometric landscape which appeared in front of me was remarkable and in many cases beautiful. I also saw that it could potentially be harmful and that much uncharted territory remains. It has been an inspiring journey and the story has just begun.

9 Sammanfattning på svenska

Bibliometri har definierats som appliceringen av matematiska och statistiska metoder på böcker och andra kommunikationsmedia. Bibliometriska metoder har genom åren använts i en rad olika sammanhang. Främst för att beskriva och bedöma forskning. Vidare har indikatorer som syftar till att ge en bild av hur stort inflytande ett universitet eller en forskargrupp har på vetenskapssamhället utvecklats. Beräkningarna är oftast baserade på hur ofta forskare runt om i världen refererar till universitetet eller forskargruppen i fråga. Hur ofta en grupp citeras av andra har i vissa sammanhang likställts med ”forskningskvalitet”. Ytterligare bibliometriska indikatorer används för att beskriva utvecklingen av ämnesområden (genom att räkna antalet publikationer inom dessa fält) eller samarbeten mellan individer eller organisationer (genom beräkningar av hur ofta dessa skriver vetenskapliga artiklar tillsammans). Den ökade användningen av bibliometri grundar sig i att metoderna anses vara objektiva och kvantitativa samt att mätningarna i sig själva inte stör forskningsprocessen. Dock har bibliometrins värde ifrågasatts eftersom metoderna sällan blivit ordentligt utvärderade.

Syftet med denna avhandling är att utforska och utveckla användbarheten av bibliometri som verktyg för att göra bedömningar av forskning. Detta sker på grundval av fyra delstudier vilka belyser olika aspekter av bibliometri. De ämnen som studeras i detalj är relaterade till hur forskningsområden identifieras (*I*), samarbeten mäts (*II*) och framgångsrik forskning kan bedömas (*I/III/IV*).

Analysen baseras primärt på data från de citeringsindex (CI) som produceras av Thomson Scientific. I de olika delstudierna beräknas bibliometriska indikatorer baserat på 24 223 (*I*), 62 104 (*II*), 6 142 055 (*III*) respektive 596 (*IV*) publikationer. För att utvärdera validiteten hos bibliometriska indikatorer baserade på CI-data kombineras dessa indikatorer med data från PubMed (*I*) och jämförs med data från manuell granskning (*I*), finansiell information (*II*), och ett svenskt system för identifiering och tidig utvärdering av nya metoder i sjukvården (*IV*). Tre nya indikatorer utvecklas också baserat på teoretiska resonemang.

Vanliga bibliometriska metoder tillåter inte säkra uppskattningar av forsknings-områdens utveckling (*I*). Bedömningar av samarbeten mellan akademi och industri baserat på samförfattade artiklar är inte heller heltäckande (*II*). Det finns också begränsningar relaterade till den nuvarande state-of-the-art indikatorn för identifiering av framgångsrik forskning (*III*). Samtidigt finns bibliometriska metoder tillgängliga som kan användas för identifiering av forskningsområden (*I*) och för att beskriva specifika typer av samarbeten mellan akademi och industri (*II*). Nya indikatorer på framgångsrik forskning som hanterar några av de problem som identifierats hos befintliga indikatorer har utvecklats (*III*). Publikationer relaterade till sjukvårdsteknologier, vilka av klinisk expertis bedöms som lovande, erhåller mycket höga mätvärden med dessa indikatorer (*IV*).

Resultaten visar att bedömningar av forskningsområden riskerar att bli miss-bedömningar om identifieringen av områdena sker baserat på rudimentära metoder. Om exempelvis tidskriftsområden används för att bestämma artiklar inom området så utelämnas samtliga artiklar i de multidisciplinära tidskrifter där utomordentliga forskningsresultat ofta publiceras. Därmed ges felaktig information som underlag för beslutsfattande. Detsamma gäller analyser av forskningssamarbeten med utgångs-punkt i samförfattarskap eftersom det finns exempel på samarbeten som inte leder till samförfattade artiklar och artiklar som är samförfattade utan att något faktiskt samarbete har ägt rum. Vidare bör noteras att bibliometriska indikatorer, vid korrekt användning, kan bidra med värdefull bakgrundsinformation till beslutsfattare. Detta gäller såväl utvecklingen av forskningsområden som forskningssamarbeten. Slutligen kan påverkan på forskarsamhället ses som en första indikator på framtida nytta för patienter och samhället.

10 Acknowledgements

Writing a thesis for doctoral degree is seldom easy. Sometimes it even seems impossible. But I have been blessed with supporters and collaborators that have made the project both possible and enjoyable. They have all contributed in their own unique way to this thesis. I would in particular like to thank:

My supervisors and directors

Professor John Skår – supervisor and former director at the Centre for Medical Innovations (CMI). For inspiring me to be my own manager and pushing me towards becoming independent.

Professor Mats Brommels, co-supervisor and director of the Medical Management Centre (MMC). For your pragmatism and efficiency without it would not have been possible to do so much in so little time.

Docent Inger Lundkvist, co-supervisor and colleague at Karolinska Institutet Strategy and Development Office (SDO). For your support and belief in that what I do is good.

Professor Göran Tomson, co-supervisor and friend. For turning the joy and art of supervision into practice. I am glad we played on the same team in this project.

Professor Maria Anvret, director of SDO. For granting me the time to finalise this thesis.

Professor Uno Forss, prefekt at the Department of Learning, Informatics, Management and Ethics (LIME). For providing an environment where strange research projects such as this can flourish.

Professor Hans Wigzell, former president of Karolinska Institutet (KI) and director of CMI. For your genuine interest in the mysterious and for creating a creative chaos.

Professor Harriet Wallberg-Henriksson, president of KI. For your endorsement and the trust shown when letting me carry out the KI Research Map 2005 project. To develop internal competence and think for oneself is a road seldom taken.

Professor Jan Carlstedt-Duke, dean of research at KI. For all the support and collaboration in developing bibliometrics into a research management tool at KI.

My friends and colleagues at Karolinska Institutet.

Carl Johan. For you energy and for your generosity with contacts and ideas.

Pia. For your administrative and parental support.

Agneta M. For being sceptical and an important scientific role model.

Agneta S. For being the only one but me who truly thinks that a report on third stream indicators is ‘fun’.

Johan A, **Maria W, Ola, Karin**, **Benjamin**, **Jocke**, **PEJ**, **Tessi, Lena**, **Anna**, **Björn** and **Emil.** For your continuous support and all memorable retreats and dinners.

Johan T. For being my opposite as a researcher -calm, thoughtful and systematic. Your considerate comments have contributed deeply to the development of this thesis.

Carl S. For sharing your room during this last year. You inspire me to improve and have taught me that there is more than Maslow.

Jesper O. For joint ventures such as one hour papers and one person randomized control trials.

Kiku. For teaching me that women do not simply complain more.

David, **Soki**, **Emma**, **Stina** and **all other PhD students at MMC.** For giving me hope for the future of medical management research and practice.

The Unit for Bioentrepreneurship. **Pauline, Theresa, Bosse, Charlotta, Hanna, Vanessa, Carl Johan and Lena.** For making the last months at LIME so much fun and for bringing a new vitality to the fifth floor in BZ!

Pelle. For being different from all other managers I've seen. The honest love you have shown me and other people in your surrounding is just amazing. Please come back soon and often!

Kinna and **Uffe**. For your encouraging cheers and all your accurate and intelligent comments on my plans and papers.

John, **Micke**, **Karin**, **Richard** and **Anders W.** For the technical development of the bibliometric system at KI which has been vital for the analyses in this thesis project**.**

Calle J. For introducing me to bibliometrics and for sharing dreams over a Staropramen.

Nina. For being so much better and still sharing and caring.

Alex and **Bita.** For all your smiles and patience with my idiotic stories.

Gustaf. For believing in elitism and defining excellence. **Björn**, **Anna**, **Anna**, **Fredrik**, **Cat**. For giving me insights into new parts of KI and supporting the faculty pub.

Mathias S. For minion and mundo.

ADG. For defining oncology.

Dr Aly. For putting the finger on the spot and inspiring me to ski Vasaloppet. Next year.

Dr Holmer. For your unique humour and wisdom.

Micke. For being a true friend and golden standard.

Friends and family

Matthias and **Annica Hulthenheim**. For making life sweeter. A special acknowledgement to **Kent** who manages the kitchen.

The GBG Crew. **Bore, Jens, Jonas, Jocke, Anna-Maria, Linda and Susanne**. For not giving up on me even though I do things as move to Stockholm and get a PhD. Life is good with you.

Johan Lundberg, my dear cousin. For the Sunday dinners with political debates and for your sincere friendship.

Gunilla. For your interest and encouraging discussions.

Jenny and **Anders.** For always having your door open and inviting me into your family.

Mari. For being genuine and for teaching how it can be easier to live with a Lundberg.

Per. For your tranquillity and generosity.

Linus, **Matilda**, **Viktor**, **Emma**, **Wilma, Hugo** and **Herman.** For always making me laugh and for bringing me hope for the future.

Sigrid. For your pride in your grandchildren.

Elsa. Tack för ditt stöd Farmor!

Pappa. For feeling good. You are an excellent teacher and competing with you has made me sharper.

Mamma. For caring and for your endless curiosity in other people.

Stefan. My wonderful brother. For your strength and courage in life. All the discussions with you have advanced my argumentation skills and made me into who I am.

Åsa. My beloved sister. For always being there. You are a great role model and inspiration.

Karolin. My amazing wife. For the love that you bring into my life everyday. For being my most critical friend and true supporter. For making me better. This thesis is for you.

11 References

1. Pritchard A. Statistical bibliography or bibliometrics? *Journal of Documentation.* 1969;25:348-349.
2. Nalimov VV, Mulchenko ZM. *Naukometriya. Izuchenie Razvitiya Nauki kak Informatsionnogo Protsessa (Scientometrics. Study of the Development of Science as an Information Process)*. English translation 1971 ed. Moscow; 1969.
3. Glänzel W. Bibliometrics as a research field: A course on theory and application of bibliometric indicators - Course Handouts 2003. http://www.norslis.net/2004/Bib_Module_KUL.pdf. Accessed January 3, 2006.
4. Lotka AJ. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science.* 1926;16(2):317-323.
5. Gross PLK, Gross EM. College libraries and chemical education. *Science.* 27 October 1927;66:385-389.
6. Price DJdS. *Little science, big science*. New York: Columbia University Press; 1963.
7. Merton RK. Science and technology in a democratic order. *Journal of Legal and Political Sociology.* 1942;1:115-126.
8. Merton RK. Priorities in scientific discovery. *American Sociological Review.* 1957;22(6):635-659.
9. Merton RK. Singletons and multiples in scientific discovery. *Proceedings of the American Philosophical Society.* 1961;105(5):470-486.
10. Garfield E. Citation Indexes for Science - New Dimension in Documentation through Association of Ideas. *Science.* 1955;122(3159):108-111.
11. Garfield E. Is Citation Analysis a Legitimate Evaluation Tool. *Scientometrics.* 1979;1(4):359-375.
12. Garfield E. The citation index - a new dimension in indexing. *Science.* 1964;144:649-654.
13. Small H. Cocitation in Scientific Literature - New Measure of Relationship between 2 Documents. *Journal of the American Society for Information Science.* 1973;24(4):265-269.
14. Debbeaver D, Rosen R. Studies in Scientific Collaboration .1. Professional Origins of Scientific Co-Authorship. *Scientometrics.* 1978;1(1):65-84.
15. Moravcsik MJ. Phenomenology and Models of Growth of Science. *Research Policy.* 1975;4(1):80-86.
16. Gilbert GN. Measuring the growth of science - review of indicators of scientific growth. *Scientometrics.* 1978;1:9-34.
17. Vlachy J. Mobility in Science - Bibliography of Scientific Career Migration, Field Mobility, International Academic Circulation and Brain-Drain. *Scientometrics.* 1979;1(2):201-228.
18. Narin F, Computer horizons inc. *Evaluative bibliometrics : the use of publication and citation analysis in the evaluation of scientific activity*. Washington, D.C.,; 1976.
19. Narin F. Objectivity Versus Relevance in Studies of Scientific Advance. *Scientometrics.* 1978;1(1):35-41.
20. Callon M, Courtial JP, Turner WA, Bauin S. From Translations to Problematic Networks - an Introduction to Co-Word Analysis. *Social Science Information Sur Les Sciences Sociales.* 1983;22(2):191-235.
21. Callon M, Courtial JP, Turner WA, Bauin S. From Translation to Network - the Co-Word Analysis. *Scientometrics.* 1983;5(1):78-78.
22. Rip A, Courtial JP. Co-Word Maps of Biotechnology - an Example of Cognitive Scientometrics. *Scientometrics.* 1984;6(6):381-400.
23. Koenig MED. Bibliometric Indicators Versus Expert Opinion in Assessing Research Performance. *Journal of the American Society for Information Science.* 1983;34(2):136-145.
24. Martin BR, Irvine J. Assessing basic research : Some partial indicators of scientific progress in radio astronomy. *Research Policy.* 1983/4 1983;12(2):61-90.

25. Noyons ECM. *Bibliometric Mapping as a Science Policy and Research Management Tool* [Doctoral Thesis]. Leiden: Centre for Science and Technology Studies (CWTS), Leiden University; 1999.
26. Kuhn TS. Historical Structure of Scientific Discovery. *Science.* 1962;136(3518):760.
27. Moed HF. *The use of bibliometric indicators for the assessment of research performance in the natural and life sciences : aspects of data collection, reliability, validity and applicability*. Leiden,; 1989.
28. van Leeuwen TN. *Second generation bibliometric indicators* [Dissertation]. Leiden: Centre for Science and Technology Studies, Leiden University; 2004.
29. Garfield E. Can citation indexing be automated? Paper presented at: Statistical association methods for mechanized documentation, 1964; Washington.
30. Duncan EB, Anderson FD, McAleese R. Qualified citation indexing: Its relevance to educational technology. In: Duncan EB, McAleese R, eds. *Information retrieval in educational technology. Proceedings of the First Symposium on Information Retrieval in Educational Technology*. Aberdeen, Scotland: University of Aberdeen; 1981:70-79.
31. Shadish WR, Tolliver D, Gray M, Sengupta SK. Author Judgments About Works They Cite - 3 Studies from Psychology Journals. *Social Studies of Science.* AUG 1995;25(3):477-498.
32. Case DO, Higgins GM. How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science.* 2000;51(7):635-645.
33. Chubin DE, Moitra SD. Content-Analysis of References - Adjunct or Alternative to Citation Counting. *Social Studies of Science.* 1975;5(4):423-441.
34. Oppenheim C, Renn SP. Highly Cited Old Papers and Reasons Why They Continue to Be Cited. *Journal of the American Society for Information Science.* 1978;29(5):225-231.
35. Moravcsik MJ, Murugesan P. Some Results on Function and Quality of Citations. *Social Studies of Science.* 1975;5(1):86-92.
36. Nicolaisen J. The J-shaped distribution of citedness. *Journal of Documentation.* 2002;58(4):383-395.
37. Gilbert GN. Referencing as persuasion. *Social Studies of Science.* 1977;7:113-122.
38. Weingart P. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics.* 2005/01// 2005;62(1):117-131.
39. Luukkonen T. Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics.* Jan 1997;38(1):27-37.
40. Moed HF. *Citation analysis in research evaluation*. Dordrecht: Springer; 2005.
41. Cozzens SE. What do citations count? The Rethoric First model. *Scientometrics.* 1989;15:437-447.
42. Latour B. *Science in action : how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press; 1987.
43. Merton RK. The sociology of science: an episodic memoir. In: Merton RK, Gaston J, eds. *The sociology of science in Europe*. Carbondale: Southern Illinois University Press; 1977:3-141.
44. Cole S, Cole JR. Scientific Output and Recognition - Study in Operation of Reward System in Science. *American Sociological Review.* 1967;32(3):377-390.
45. Oxford English dictionary [Oxford]: Oxford University Press; 2000.
46. Hand DJ. *Measurement theory and practice: the world through quantification*. London: Arnold; 2004.
47. Sentinel Event Glossary of Terms. http://www.jointcommission.org/SentinelEvents/se_glossary.htm. Accessed Sept 27th, 2006.
48. van Raan AFJ. Measuring Science - Capita Selecta of Current Main Issues. In: Glänzel W, Schmoch U, Moed HF, eds. *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems*. Dordrecht ; London: Kluwer Academic Publishers; 2004:19-50.
49. Grupp H, Mogee ME. Indicators for national science and technology policy. In: Glänzel W, Schmoch U, Moed HF, eds. *Handbook of quantitative science and*

technology research : the use of publication and patent statistics in studies of S&T systems. Dordrecht ; London: Kluwer Academic Publishers; 2004:75-94.
50. Research Evaluation and Policy Project. *Quantitative indicators for research assessment – a literature review.* Canberra: The Australian National University; 2005. REPP Discussion Paper 05/1.
51. Garfield E. Citation Analysis as a Tool in Journal Evaluation - Journals Can Be Ranked by Frequency and Impact of Citations for Science Policy Studies. *Science.* 1972;178(4060):471-+.
52. Garfield E. *Citation indexing : its theory and application in science, technology, and humanities*. New York: Wiley; 1979.
53. H.F. Moed TNVL. Improving the accuracy of institute for scientific information's journal impact factors. *Journal of the American Society for Information Science.* 1995;46(6):461-467.
54. Garfield E. Fortnightly Review: How can impact factors be improved? *British Medical Journal.* August 17, 1996 1996;313(7054):411-413.
55. Opthof T. Sense and nonsense about the impact factor. *Cardiovasc Res.* Jan 1997;33(1):1-7.
56. Seglen PO. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal.* Feb 15 1997;314(7079):498-502.
57. Hecht F, Hecht BK, Sandberg AA. The journal "impact factor": a misnamed, misleading, misused measure. *Cancer Genet Cytogenet.* Jul 15 1998;104(2):77-81.
58. Bloch S, Walter G. The Impact Factor: time for change. *Aust N Z J Psychiatry.* Oct 2001;35(5):563-568.
59. Glänzel W, Moed HF. Journal impact measures in bibliometric research. *Scientometrics.* FEB 2002;53(2):171-193.
60. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? *J Med Libr Assoc.* Jan 2003;91(1):42-46.
61. Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America.* Nov 15 2005;102(46):16569-16572.
62. Hirsch JE. An index to quantify an individual's scientific research output. http://xxx.arxiv.org/abs/physics/0508025. Accessed 2005-12-29, 2005.
63. Bornmann L, Daniel HD. Does the h-index for ranking of scientists really work? *Scientometrics.* DEC 2005;65(3):391-392.
64. van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups: a Manuscript. 2005:12, Leiden.
65. Schubert A, Glänzel W, Braun T. Relative citation rate: A new indicator for measuring the impact of publications. Paper presented at: First National Conference with International Participation on Scientometrics and Linguistics of Scientific Text, 1983; Varna.
66. Vinkler P. Evaluation of Some Methods for the Relative Assessment of Scientific Publications. *Scientometrics.* Sep 1986;10(3-4):157-177.
67. Moed HF, Debruin RE, Vanleeuwen TN. New Bibliometric Tools for the Assessment of National Research Performance - Database Description, Overview of Indicators and First Applications. *Scientometrics.* JUL-AUG 1995;33(3):381-422.
68. Seglen PO. The Skewness of Science. *Journal of the American Society for Information Science.* Oct 1992;43(9):628-638.
69. Laudel G. What do we measure by co-authorships? *Research Evaluation.* Apr 2002;11(1):3-15.
70. Katz JS, Martin BR. What is research collaboration? *Research Policy.* MAR 1997;26(1):1-18.
71. Beaver DD, Rosen R. Studies in Scientific Collaboration .1. Professional Origins of Scientific Co-Authorship. *Scientometrics.* 1978;1(1):65-84.
72. Melin G, Persson O. Studying research collaboration using co-authorships. *Scientometrics.* JUL-AUG 1996;36(3):363-377.
73. Melin G. Globalization or internationalization in the sciences. *Sociologisk Forskning.* 1999;36(3):22-36.

74. Beaver DD. Reflections on scientific collaboration, (and its study): past, present, and future. *Scientometrics.* Nov-Dec 2001;52(3):365-377.
75. Tijssen RJW. Is the commercialisation of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles. *Research Policy.* Jul 2004;33(5):709-733.
76. Glänzel W, Schubert A. Analysing scientific networks through co-authorship. In: Glänzel W, Schmoch U, Moed HF, eds. *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems*. Dordrecht ; London: Kluwer Academic Publishers; 2004:257-276.
77. Carayannis EG, Laget P. Transatlantic innovation infrastructure networks: public-private, EU-US R&D partnerships. *R & D Management.* Jan 2004;34(1):17-31.
78. Butcher J, Jeffrey P. The use of bibliometric indicators to explore industry-academia collaboration trends over time in the field of membrane use for water treatment. *Technovation.* Nov 2005;25(11):1273-1280.
79. Wang Y, Wu YS, Pan YT, Ma Z, Rousseau R. Scientific collaboration in China as reflected in co-authorship. *Scientometrics.* Feb 2005;62(2):183-198.
80. Luukkonen T, Tijssen RJW, Persson O, Sivertsen G. The Measurement of International Scientific Collaboration. *Scientometrics.* SEP 1993;28(1):15-36.
81. Tijssen RJW, van Leeuwen TN, Korevaar JC. Scientific publication actitivy of industry in the Netherlands. *Research Evaluation.* 1996;6:105-119.
82. Tijssen RJW, van Wijk E. In search of the European Paradox: an international comparison of Europe's scientific performance and knowledge flows in information and communication technologies research. *Research Policy.* JUN 1999;28(5):519-543.
83. Meyer M, Bhattacharya S. Commonalities and differences between scholarly and technical collaboration - An exploration of co-invention and co-authorship analyses. *Scientometrics.* 2004;61(3):443-456.
84. Calero C, van Leeuwen TN, Tijssen RJW. Research Networks of Pharmaceutical Firms: Geographical Patterns of Research Collaboration within and between Firms. Paper presented at: ISSI 2005, 2005; Stockholm.
85. Subramanyam K. Bibliometric Studies of Research Collaboration - a Review. *Journal of Information Science.* 1983;6(1):33-38.
86. Mela GS, Cimmino MA. An overview of rheumatological research in the European Union. *Annals of the Rheumatic Diseases.* Nov 1998;57(11):643-647.
87. Mela GS, Cimmino MA, Ugolini D. Impact assessment of oncology research in the European Union. *Eur J Cancer.* Aug 1999;35(8):1182-1186.
88. Garcia-Rio F, Serrano S, Dorgham A, et al. A bibliometric evaluation of European Union research of the respiratory system from 1987-1998. *Eur Respir J.* Jun 2001;17(6):1175-1180.
89. Ugolini D, Mela GS. Oncological research overview in the European Union. A 5-year survey. *European Journal of Cancer.* 2003/9 2003;39(13):1888-1894.
90. Grossi F, Belvedere O, Rosso R. Geography of clinical cancer research publications from 1995 to 1999. *Eur J Cancer.* Jan 2003;39(1):106-111.
91. Ugolini D, Cimmino MA, Casilli C, Mela GS. How the European Union writes about ophthalmology. *Scientometrics.* Sep 2001;52(1):45-58.
92. Davis M, Wilson CS. Research contributions in ophthalmology: Australia's productivity. *Clinical and Experimental Ophthalmology.* Aug 2003;31(4):286-293.
93. Belinchon I, Ramos JM, Sanchez-Yus E, Betlloch I. Dermatological scientific production from European Union authors (1987-2000). *Scientometrics.* 2004;61(2):271-281.
94. Skram U, Larsen B, Ingwersen P, Viby-Mogensen J. Scandinavian research in anaesthesiology 1981-2000: visibility and impact in EU and world context. *Acta Anaesthesiologica Scandinavica.* Sep 2004;48(8):1006-1013.
95. Noyons ECM, Moed HF, Luwel M. Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science.* FEB 1999;50(2):115-131.

96. Mohanti BK, Bansal M, Sharma DN, Gairola M, Majhail NS. Publications by radiation oncologists from India: A brief survey. *National Medical Journal of India.* Jul-Aug 2004;17(4):197-199.
97. Adams J, Bailry T, Jackson L, Scott P, Pendlebury D, Small H. Benchmaring of the international standing of research in England - Report of a consultancy study on bibliometric analysis. November 1997 1997:174.
98. Butler L, Biglia B, Bourke P. *Australian Biomedical Research: Funding acknowledgments and performance* December 1998 1998.
99. Lewison G. The definition of biomedical research subfields with title keywords and application to the analysis of research outputs. *Research Evaluation.* 1996;6(25):25-36.
100. Ugolini D, Casilli C, Mela GS. Assessing oncological productivity: is one method sufficient? *European Journal of Cancer.* 2002/5 2002;38(8):1121-1125.
101. U.S. National Library of Medicine. Medical Subject Headings. Website]. 2004-12-13; http://www.nlm.nih.gov/mesh/meshhome.html. Accessed Jan 03, 2005.
102. Lindsey D. The Corrected Quality Ratio: A Composite Index of Scientific Contribution to Knowledge. *Social Studies of Science.* 1978;8(3):349-354.
103. van Raan AFJ. Measurement of Central Aspects of Scientific Research: Performance, Interdisciplinarity, Structure. *Measurement: Interdisciplinary Research and Perspectives.* 2005;3(1):1-19.
104. van Leeuwen TN, Visser MS, Moed HF, Nederhof TJ, van Raan AFJ. Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics.* 2003;57(2):257-280.
105. Karlsson S, Wadskog D. *Hur mycket citeras svenska publikationer? Bibliometrisk översikt över Sveriges vetenskapliga publicering mellan 1982 och 2004.* Stockholm: Vetenskapsrådet; 2006.
106. van Raan AFJ. The influence of international collaboration on the impact of research results - Some simple mathematical considerations concerning the role of self-citations. *Scientometrics.* JUL-AUG 1998;42(3):423-428.
107. Persson O, Glänzel W, Danell R. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics.* 2004;60(3):421-432.
108. Glänzel W, Schoepflin U. A Bibliometric Study on Aging and Reception Processes of Scientific Literature. *Journal of Information Science.* 1995;21(1):37-53.
109. Luwel M, Moed HF. Publication delays in the science field and their relationship to the ageing of scientific literature. *Scientometrics.* JAN-FEB 1998;41(1-2):29-40.
110. van Raan AFJ. Reply to the comments of Liu et al. *Scientometrics.* July 2005;64(1):111-112.
111. International Working Party to Promote and Revitalise Academic Medicine. Academic medicine: the evidence base. *British Medical Journal.* Oct 2 2004;329(7469):789-792.
112. Narin F, Hamilton KS. Bibliometric performance measures. *Scientometrics.* Jul-Aug 1996;36(3):293-310.
113. Raymond C. Impact factors: Use and abuse in biomedical research. *The Anatomical Record.* 1999;257(2):54-57.
114. Roessner JD. Outcome measurement in the USA: state of the art. *Research Evaluation.* 1 August 2002;11(2):85-93.
115. Clark KE. The APA study of psychologists. *American Psychologist.* 1957;9:117-120.
116. Butler L. What happens when funding is linked to publication counts? In: Glänzel W, Schmoch U, Moed HF, eds. *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems*. Dordrecht ; London: Kluwer Academic Publishers; 2004:389-405.
117. National Library of Medicine. PubMed Overview. December 29. Accessed Jan 4 2006.
118. Fleiss JL, Levin BA. *Statistical methods for rates and proportions*. 2. ed. New York: Wiley; 1981.

119. Pezzullo JC. 2-way Contingency Table Analysis. Website]. Jan 5 2005; http://members.aol.com/johnp71/ctab2x2.html. Accessed April 27 2005, 2005.
120. Heffner AG. Funded Research, Multiple Authorship, and Sub-Authorship Collaboration in 4 Disciplines. *Scientometrics.* 1981;3(1):5-12.
121. Lundberg J, Fransson A, Brommels M, Skar J, Lundkvist I. Is it better or just the same? Article identification strategies impact bibliometric assessments. *Scientometrics.* 2006;66(1):183-197.
122. Lundberg J, Tomson G, Lundkvist I, Skar J, Brommels M. Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics.* 2006;69(3):In press.
123. Lundberg J. Lifting the crown - Citation z-score. *Journal of Informetrics.* 2006;Accepted.
124. Kirkwood BR, Sterne JAC. *Essential medical statistics*. 2. ed. Malden, Mass.: Blackwell Science; 2003.
125. Lundberg J, Brommels M, Skar J, Tomson G. Breakthroughs and bibliometrics: The sources of the Swedish rapid evaluation system of technologies in health care. *Submitted.* 2006.
126. Bland M. *An introduction to medical statistics*. 3. ed. Oxford: Oxford University Press; 2000.
127. Lundberg J, Tomson G, Lundkvist I, Skar J, Brommels M. Collaboration uncovered: Exploring the adequacy to measure university-industry collaboration through co-authorship and funding. *Submitted.* 2005.
128. Tijssen RJW, Visser MS, van Leeuwen TN. Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics.* 2002;54(3):381-397.
129. Øvretveit J. A Framework for Quality Improvement Translation: Understanding the Conditionality of Interventions. *Joint Commission Journal on Quality and Patient Safety.* 2004;30(Global Supplement):15-24.
130. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *British Medical Journal.* Apr 22 2000;320(7242):1107-1111.
131. van Leeuwen TN. Descriptive versus Evaluative Bibliometrics. In: Glänzel W, Schmoch U, Moed HF, eds. *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems*. Dordrecht ; London: Kluwer Academic Publishers; 2004:373-388.
132. Taylor J. The Impact of Performance Indicators on the Work of University Academics: Evidence from Australian Universities. *Higher Education Quarterly.* 2001;55(1):42-61.
133. Brenner S. Loose end. *Current Biology.* 1995;5(5):568-568.