CITATIONS AND THEIR USE AS INDICATORS IN SCIENCE POLICY. STUDIES OF VALIDITY AND APPLICABILITY ISSUES WITH A PARTICULAR FOCUS ON HIGHLY CITED PAPERS

DISSERTATION

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof.dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday 31st March 2005 at 16.45

by

Dag Westreng Aksnes

born on 6th June 1969
in Bergen, Norway

This dissertation has been approved by the promotor:
prof. dr. Arie Rip
School of Business, Public Administration and Technology
Centre for Studies of Science, Technology and Society

Citations and their use as indicators in science policy

Studies of validity and applicability issues with a particular focus on highly cited papers

Dag W. Aksnes

Dissertation for the doctoral degree of the University of Twente

March 2005

Table of contents

Preface

The studies upon which this PhD dissertation is based have been carried out at the Norwegian Institute of Studies in Research and Higher Education (NIFU STEP). I commenced my career here by working on R&D indicators and became involved at an early stage in bibliometrics and the production of publication and citation indicators for various reports and evaluations. When undertaking this work I was struck by the extreme skewness that characterised citation distributions. In particular, I was interested as to why this skewness arose and what implications it had for the use of citations as indicators. These questions also influenced my choice of research topics for this dissertation.

More generally, through the choice of research topics I wished to learn more about the possibilities and limitations of bibliometric indicators, since such knowledge is important in my daily work as a researcher at NIFU STEP. I must admit, however, that I have always been ambivalent towards bibliometrics as a field of study. On the one hand, publication and citation analysis can yield much interesting information on various issues concerning science and scientific communication. On the other, there are many possibilities of misuse of such data in the context of research management. In fact, it is my intention that this thesis will contribute to highlight the duality of potentials and dangers inherent in bibliometric analysis.

The studies have been carried out on part-time basis. Although I did not receive a PhD fellowship to conduct this research, various sources of funding have enabled me to obtain time for working on the thesis. Some of the work was made possible through a strategic institute program on the output of R&D funded by the Research Council of Norway. NIFU STEP has also supported my work by allowing me to work on the PhD on part-time basis.

For some time I was looking for a supervisor in various Norwegian university departments, but it turned out that it was difficult to find a person with sufficient competence within this research field. Further, the project did not fit thematically into the Norwegian PhD system based on traditional academic disciplines. I was honoured when I later was accepted as a PhD student by professor Arie Rip as part of NIFU STEP's appointment of collaboration with the Centre for Studies of Science, Technology and Society at The University of Twente. It is unfortunate however, that I did not have the

opportunity to stay at University of Twente for a longer period. Except for occasional meetings for supervision, I have been a ‘distant student’.

There are many people that I wish to thank for their assistance and guidance. First of all, my supervisor Arie Rip. I have learned a lot from Arie’s supervision and thoughtful advices, his penetrating comments and valuable suggestions. Moreover, I am grateful to Gunnar Sivertsen for his support and help in the design of the project. I would also like to thank co-authors of certain articles: Terje Bruen Olsen, Per O. Seglen, and Randi E. Taxt for their valuable contributions. Furthermore, I wish to express my gratitude to NIFU STEP for providing me with the opportunity to carry out this project. Without this support the project would never have been realised.

Finally, a special thank to my wife Kristin for her love, care and support in my endeavours.

Oslo, January 2005

Dag W. Aksnes

1. Introduction to citations and their use as indicators

Citations have increasingly been applied as performance indicators in the context of science policy and research evaluation. The basic assumption underlying such applications is that number of citations can be regarded as a measure of scientific quality or impact. Over the years a large number of studies have been carried out to ascertain the extent to which this assumption is fulfilled. Many studies have also found that citation indicators correspond fairly well, especially in the aggregate, with various measures of research performance or scientific recognition which are taken as reflecting quality. On the other hand, there have been several studies challenging or criticising such use of citations. Particularly the application of citation indicators for evaluative purposes has been surrounded by controversy. Thus, the question of what is measured by citations and the validity and reliability of citation indicators is still much debated.

This dissertation aims at contributing towards the discussion on the use of citations as indicators. A particular focus is directed towards highly cited papers. Because citation distributions are extremely skewed in which most publications are poorly cited or not cited at all and a few publications are very highly cited, it is clear that this phenomenon has to be taken into account when constructing and using citation indicators. It is thus of particular interest to analyse patterns of highly cited papers. This then leads on to a number of studies of the methodological basis and validity of citation indicators.

The dissertation consists of a collection of journal articles, preceded by two introductory chapters. A variety of different topics are addressed, all concerning citations and citation indicators. The topics selected are questions I have regarded as important and that need further attention. In other words, I have focussed on issues that given the present level of knowledge are interesting and urgent – even if they do not cover all such issues. There have also been practical limitations: the research would have to be undertaken within the framework of a PhD thesis and discuss topics where necessary data is available. All the studies are based on publications and citation data relating to Norwegian science.

Chapter 1 gives a brief overview of citations and their use as indicators. In Chapter 2 a preview of the studies and their findings is given, together with general remarks on data and methodology. Chapters 3–9 contain the seven journal articles of the dissertation. Five of the articles have previously been published, one has been accepted for publication, and one will be submitted for publication. There is some overlap between the articles in terms of the literature discussed, while Chapters 1 and 2 also anticipate (sometimes literally) on the introductions and conclusions of the various articles.

1.1 Citations, citation indexes, and citation indicators

1.1.1 Citations and their roles

Citations represent an important component of scientific communication. Already prior to the 19th century it was a convention that scientists referred to earlier literature relating to the theme of the study (Egghe & Rousseau, 1990: 204). The references are intended to identify earlier contributions (concepts, methods, theory, empirical findings, etc.) upon which the present contribution was built, and against which it positions itself. Thus, it is a basic feature of the scientific article that it contains a number of such references and that these references are attached to specific points in the text.

The terms ’reference’ and ’citation’ are often used interchangeably. However, there is a distinction – as de Solla Price already emphasised: ”…if Paper R contains a bibliographic footnote using and describing Paper C, then R contains a *reference* to C, and C has a *citation* from R.” (Price, 1970).[1]

Citations have been the loci for a large number of studies, for example of intellectual patterns and historical knowledge lineages. The spread of citation analysis depends on the availability of citation indexes, in particular ISI’s *Science Citation Index*. It is, however, not a new phenomenon. In the second half of the 19th century Frank Shepard of the United States created a citation index covering judicial decisions (Wouters, 1999: 22). This index was used as a tool for lawyers to see whether a legal procedure was still valid. Moreover, some research librarians have systematically applied citation analyses since the early years of the last century. Among the first examples of such applications are papers by J. R. Cole and N. B. Eales (1917) and P. L. K. Gross and E. M Gross (1927) (Wouters, 1999: 4). A main idea in these studies was to assemble data on how frequently

[1] In this thesis I have only to a certain extent managed to adopt this distinction. While I have attempted to use the term ‘reference’ in an unambiguous way, ‘citation’ sometimes is used in both meanings. This is because of the common practice to use this word in both ways. For example, expressions such as the ‘citation process’ have often been used (Cronin, 1984) while it would have been more appropriate to speak of the process of inserting references.

journals were cited. In turn, these data were considered to indicate the value of subscriptions to the journals.

It was not, however, until the creation of the *Science Citation Index* (SCI) by Eugene Garfield in 1961 that citation analysis really emerged as a separate field of study. This bibliographic database was originally developed for information retrieval purposes, to aid researchers in locating papers of interest in the vast research literature archives (Welljams-Dorof, 1997). As a subsidiary property it enabled scientific literature to be analysed quantitatively. Since the 1960s the *Science Citation Index* and similar bibliographic databases located at the Institute for Scientific Information (ISI) have been applied in a large number of studies and in a variety of fields. The possibility for citation analyses has been an important reason for this popularity. As part of the indexing process, ISI systematically registers all the references of the indexed publications. These references are organised according to the publications they point to. On this basis each publication can be attributed a citation count showing how many times each paper has been cited by later publications indexed in the database. Citation counts can then be calculated for aggregated publications representing, for example, research units, departments, or scientific fields.

Because citations may be regarded as the mirror images of the references, the use of citations as indicators of research performance needs to be justified or grounded in the referencing behaviour of the scientists (Wouters, 1999). If scientists cite the work they find useful, frequently cited papers are assumed to have been more useful than publications which are hardly cited at all, and possibly be more useful and thus important in their own right. Thus, the number of citations may be regarded as a measure of the article's usefulness, impact, or influence. The same reasoning can be used for aggregated levels of articles. The more citations they draw, the greater their influence must be. Robert K. Merton has provided the original theoretical basis for this link between citations and the use and quality of scientific contribution. In Merton's traditional account of science, the norms of science oblige researchers to cite the work upon which they draw, and in this way acknowledge or credit contributions by others (Merton, 1979). Such norms are upheld through informal interaction in scientific communities and through peer review of manuscripts submitted to scientific journals.

Empirical studies have shown that the Mertonian account of the normative structure of science covers only part of the dynamics. For the citation process, this implies that other incentives occur, like the importance of creating visibility for one's work,[2] and being selective in referencing to create a distance between oneself and others. Merton himself already pointed out the ambivalence of the norms, for example that one should not hide one's results from colleagues in one's community, but also not rush into print before one's findings are robust. Merton also identified system level phenomena like the "Matthew effect": to whom who has shall be given more (see further section 1.1.2).

I shall come back to the citation process in more detail in section 1.1.2, and just note here that clearly, a work may be cited for a large number of reasons including tactical ones such as citing a journal editor's work as an attempt to enhance the chances of acceptance for publication. Whether this affects the use of citations as performance indicators is a matter of debate.

The other main entry point for citation studies has been the availability of the data provided by the Institute for Scientific Information (now owned by the Thomson Corporation and named *Thomson ISI*). In fact, the study of citations might not even have existed as a field of study if it had not been for the existence of this database located in Philadelphia in the United States. The findings of citation studies are thus dependent upon the particular characteristics of this database, how it is constructed, its coverage, and so forth.

The database covers a large number of specialised and multidisciplinary journals within the natural sciences, medicine, technology, the social sciences and the humanities. The coverage varies between the different database products. According to the website of the Thomson ISI company,[3] the most well-known product the *Science Citation Index* today covers 3,700 journals, and the expanded version of this publication database (*Science Citation Index Expanded)* 5,800 journals. The online product *Web of Science* covering the three citation indexes *Science Citation Expanded*, *Social Sciences Citation Index*, and *Arts & Humanities Citation Index* includes 8,500 journals. Compared to the large volume of scientific and scholarly journals that exist today, this represents a limited part. The selection of journals is based on a careful examination procedure in which a journal must meet particular requirements in order to be included (Testa, 1997). Even of its coverage is not complete, the ISI database will include all major journals and is generally regarded as

[2] Cf. R. Collins (1975: 480) who likened science to "an open plain with men scattered throughout it, shouting: 'Listen to me! Listen to me!'"
[3] URL: http://www.isinet.com

constituting a satisfactory representation of international mainstream research (Katz & Hicks, 1998).

From a bibliometric perspective, a main advantage of the ISI database is that it fully indexes the journals that are included. Moreover, all author names, author addresses and references are indexed. Through its construction it is also well adapted for bibliometric analysis. For example, country names and journal names are standardised, controlled terms. It is also an advantage that it is multidisciplinary in contrast to most other similar databases which cover just one or a few scientific disciplines. The bibliometric studies in this thesis are based upon ISI-data, and the further discussions will accordingly refer to this database.

At a general level the average number of citations per paper as measured by the ISI-indices has been increasing over the years (cf. the *National Science Indicators*). Not only because each paper contains more references (Garfield, 1980; Moed, Burger, Frankfort, & van Raan, 1985) but also because the number of journals (and papers) indexed by ISI has been increasing (Garfield, 1998b; Moed, Burger, Frankfurt, & van Raan, 1985).

De Solla Price showed quite early that recent papers are more cited than older ones (Price, 1965). Nevertheless, there are large individual as well as disciplinary differences. The citation counts of an article may vary from year to year. However, most articles are cited relatively constantly from year to year following a general pattern of rise and fall. In the natural sciences and medicine with only occasional exceptions articles more than 15 years old are hardly cited at all (Oppenheim & Renn, 1978). Although the peak in the number of citations received in a year differs between the scientific fields, a maximum is on the whole reached about three years after publication (Van Raan, 1993).

As described above, citation distributions are extremely skewed. This skewness was also early identified by Solla Price (Price, 1965). The large majority of the scientific papers are never or seldom cited in the subsequent scientific literature. On the other hand some papers have an extremely large number of citations. The most highly cited paper ever is an article from 1951 by O. H. Lowry *et al.* on protein measurement (1951). This article has now been cited more than 250,000 times (Garfield, 1997b).

Citation rates vary considerably between different subject areas. For example, on average papers in molecular biology contain many more references than mathematics papers (Garfield, 1979b). Accordingly, one observes a much higher citation level in molecular biology than in mathematics. Generally, the average citation rate of a scientific field is determined by different factors, most importantly the average number of references per paper. In addition, the percentage of these references that appears in ISI-indexed journals, the average age of the references, and the ratio between new publications in the field and the total number of publications, are relevant.

1.1.2 The citation process

The references in scientific articles to other publications represent the building blocks of citation indicators. Therefore, the citing behaviour of scientists has a direct relationship to the value of citation analysis (Wouters, 1999). The question of what citations “measure” must be based on the reference behaviour of scientists. For a long time, bibliometricians and sociologists of science have tried to construct such a “citation theory”. In 1981, Smith could still complain that: “Not enough is known about the ‘citation behavior’ of authors – why the author makes citations, why he makes his particular citations, and how they reflect or do not reflect his actual research and use of the literature. When more is learned about the actual norms and practices involved, we will be in a better position to know whether (and in what ways) it makes sense to use citation analysis in various application areas.” (Smith, 1981). Even now, in spite of detailed studies of referencing behaviour and of aggregate effects, there is no unified theory.

Empirical studies have revealed a multitude of motivations, functions and causes of references in scientific communication. The role of the reference, both in the citing text and with respect to the cited text have turned out to be equally varied (Wouters, 1999). Moreover, the various studies differ in methodology and in ways of classifying the citations, and because of this it is difficult to extract any common or general conclusion. The quest for a general theory has also been hampered by the fact that the studies often have been based on rather small samples of papers from selected scientific fields.

One approach is to look at the norms and conventions of the citation process as determining the basic structure and characteristics, while other features are then seen as deviations from the “ideal” situation. The occurrence of deviations would then call for better control and sanctions, for example through peer review. Although it is generally accepted that the Mertonian account of the normative structure of science does not satisfactorily cover the dynamics, there definitely are binding rules of scientific communication, in particular to cite the work one is drawing upon. For example, Davenport and Cronin (2000) emphasis that infringements to this rule may invite potentially severe sanctions. This is the basic justification for the use of citations to find substantial linkages between scientific publications. Because such rules are embedded in the reward system of science, and quality is what is being rewarded (often), citations can become indicators of quality.

There are further, less general norms. They may concern how and when to cite, and how many references a paper should contain. Some vary from discipline to discipline (in particular between the natural sciences/medicine on the one hand and the humanities and the social sciences on the other). According to Wouters (1999) it is better, therefore, to speak of citation cultures in the plural form.

Considerations of the citation behaviour of scientists have shown that the citation process is complex. In an early work Garfield (1964) suggested 15 different reasons for why authors cite other publications (reprinted in Garfield, 1977). Among these were paying homage to pioneers, giving credit for related work, identifying methodology, providing background reading, correcting a work, criticizing previous work, substantiating claims, alerts to a forthcoming work, providing leads to poorly disseminated work, authenticating data and classes of fact – physical constants, etc., identifying original publications in which an idea or concept was discussed, identifying original publication or other work describing an eponymic concept, disclaiming works of others and disputing priority claims.

Similarly, the textual function of citations may be very different. In a scientific article some of the references will represent works that are crucial or significant antecedents to the present work; others may represent more general background literature. For example, reviewing the literature published on this topic during 1965–1980, Small (1982) identified five distinctions: a cited work may be 1) refuted, 2) noted only, 3) reviewed, 4) applied, or 5) supported by the citing work. These categories were respectively characterised as: 1) negative, 2) perfunctory, 3) compared, 4) used 5) substantiated. Thus, the different functions citations may have in a text are much more complex than merely providing documentation and support for particular statements.

According to Law (1986), it is typical that the introduction of a scientific article is structured as a progression from the general to the particular. References have been found to be most frequent in the introductory section a paper (Hargens, 2000). Thus, in the introduction, an article typically refers to more general or basic works within a field. The net effect of many articles referring to the same general works, therefore, is that such contributions get a very large number of citations. This was indicated in a case study by

Voos and Dagaev (1976): References to highly cited publications seemed to occur more in the introduction than anywhere else in the articles.

Similarly, since most scientific articles contain a methodology section in which the methods applied in the study are documented, authors typically cite the basic papers describing these methods. This may explain why some papers containing commonly used methods sometimes receive a very large number of citations. The prime example here is the paper by Lowry *et al.* on protein measurement referred to above.

Others have analysed the process of how authors select references. An article may rest on a knowledge basis built up through hundreds or thousands of former publications. Only a small proportion of these publications are cited. The citation process is influenced by what the author happens to have read, and may include tactical moves such as citing the journal editor's or expected peer reviewers' work (cf. Case & Higgins, 2000). According to Gilbert, (1977) citing (”referencing”) is essentially a device for persuasion. A scientist has to persuade the scientific community of the value and importance of his work and in this respect the references function as important rhetorical tools. Because the references vary in their power of persuasion, it will be more persuasive to cite an authoritative paper. Thus, an author tends to select references that will be regarded as authoritative by the intended audience.

Compared with the traditional, so-called Mertonian approach, these analyses of the citation process focus on struggles, rhetorics, tactical and strategic games. In doing so, further complexities of the dynamics of science, including citation processes, are highlighted. What remains, however, that there are patterns which are reproduced (even while slowly evolving over longer periods of time), which depend on rules for behaviour and interaction, even if not the traditional Mertonian ones. Instead of standard ("ideal") versus deviation, the question now is to understand the pattern, and perhaps identify ways to link quality or substance to particular features of citation processes.

In the literature, the two different approaches have been contrasted as emphasizing the cognitive function of the references; versus analysing citing as a social process. Typically, the latter approach would focus on ”outside” and social factors rather than content, and has been associated with attempts to undermine the use of citations as indicators of quality. There have been extensive debates between the proponents of the two positions over the years. Still, the different approaches need not preclude each other, as I intimated already. In fact, some authors have tried to develop a multidimensional approach

(Amsterdamska & Leydesdorff, 1989; Callon, Law, & Rip, 1986; Cozzens, 1989; Luukkonen, 1997b).

Cozzens, for example, has emphasised that a pluralistic explanation of citations means that we accept aspects of all perspectives. In the course of writing a paper a scientist's actions may be oriented to one or another aspect. On the one hand the citation behaviour of individuals is affected by external pressures and there are personal motives, self interests and so forth in the citation process; on the other, there are certain norms, rules, traditions and etiquettes that limit the scope and acceptability of individual actions.

Building on this multi-dimensional approach one can inquire whether one can still disentangle the various dynamics, at least in the aggregate. This would then lead to a more sophisticated justification of using citation indicators (and perhaps further proposals to construct and use them). In Article II (Chapter 4) I have introduced the concepts of quality dynamics and visibility dynamics as a first and still somewhat programmatic attempt. The case of highly-cited papers is particularly suited to discuss the basis for such a distinction.

High citation counts appear as a result of many researchers' decisions to cite a particular paper. In order to explain how some papers come to be highly cited, one has to focus on how these micro-level processes aggregate. The conceptual distinction between quality dynamics and visibility dynamics is at least heuristically useful to find out more about these processes.

The concept of quality dynamics has its foundation in the fact that scientific progress is generated by a variety of contributions. Some represent a major scientific advance; others are filling in the details. In addition, there is a distinction between core knowledge (stabilized and shared) and frontier knowledge (Cole, 1992). According to Cole, core knowledge consists of the basic theories and findings within a field, while frontier knowledge is the knowledge currently being produced. A large part of what it is published does not make its way into core knowledge (if it does, the journey is marked by the reception of many citations). This differential fate occurs because much of the research produced at the frontier aims at producing low-level descriptive analyses or comprises contributions that turn out to be of little or no lasting significance (Cole, 2000). This view seems plausible; in addition to anecdotal evidence, there is some analysis of the full production in a scientific specialty in these terms which shows this pattern (May, 1968). A further consideration is that publications may not be cited because their research question is a "dead end" so does not serve as a positive basis for further work – even if the expansion of knowledge also requires insight in what does <u>not</u> work out. If scientists tend to cite works that are useful or important for their own research, as such or to position it in the literature, it is clear that some core publications will be cited very often, and a tiny fraction of frontier publications on their way to the core will draw a lot of citations as well. Thus, one should expect a skewed distribution of citation scores of papers. In this view, highly cited papers can be equated with major or important contributions in the core or in the frontier.

Of course, the relationship will not be unambiguous. One confounding factor is that the importance of a contribution cannot always be recognized immediately by the scientists themselves. Citations can only reflect the assessments of the scientists at the time, while contributions can often only be properly assessed after some time has passed. It is for this reason that visibility dynamics cannot be simply separated from quality dynamics. Citation patterns cannot be interpreted as a reflection of quality dynamics. In one of the studies of this dissertation (Article VI), it is shown that highly cited articles are not necessarily considered by the authors themselves as major contributions (although the large majority are).

Visibility dynamics, while entangled with quality dynamics, do have some characteristics of their own. Particularly important is the bandwagon effect: When a paper is cited in many subsequent papers, even more people become aware of this paper. Thus, its visibility increases and thereby the chances of getting even more citations. This is also a coping strategy for individual researchers having difficulty to keep up with the growing literature: Cite what other people cite, because this is first-round indication that it is “signal” rather than “noise”. Thus, there will be a bias in favour of previously highly cited papers.

Another mechanism but with the same result is known as the “Matthew effect” (Merton, 1968), a biblical reference to the Book of Matthew’s phrase about the rich getting richer and the poor getting poorer.[4] Robert K. Merton introduced this concept referring to a phenomenon that in science as in other areas of human life, those who have are likely to be given more (Merton, 1968, 1988). Taking inspiration from Harriet Zuckerman’s interviews with Nobel laureates, Merton argued that recognition is skewed in favour of established scientists: “The ‘Matthew effect’ consists of accruing greater increments of recognition for particular scientific contributions by scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark.” (Merton, 1988).

[4] Matthew 13:12: “Whoever has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him.”

With respect to citations this means that articles written by famous scientists are likely to receive more citations then they would if written by relatively unknown scientists. For example, R. C. Lewontin and J. L. Hubby argued this was the case when comparing the citation counts of two of their collaborative papers (see Merton, 1988). While this may occur, the evidence is not unambiguous, also because scientists are aware of the Matthew effect and might go out of their way (if they can spare the time) to balance their references.

It has been shown that significant works tend to be utilized regardless of who has written them and do not need a “Matthew effect” to attain visibility (Cole & Cole, 1973). Many highly cited papers have been written by scientists who were relatively unknown at the time of their writing. It is also typical that articles written by even the most recognised authors show large variability in citedness (Seglen, 1994). This indicates that citation frequencies are determined more by content than by authors. What the Matthew effect still might do is to enlarge the differences between the already known and the unknown scientists, making the citation distribution more skewed.

A similar dynamics, but derived from rhetorical force rather than reputation per se, has been identified by Gilbert discussing referencing as persuasion (Gilbert, 1977). When an article begins to acquire visibility, it will become more valuable as a reference. This further increases the number of citations. If the spiralling continues, a few exemplary papers will receive very high citation rates.

Thus, there is also a temporal dimension to the citation process. An article may first be cited for substantial reasons (e.g., its content has been used). Later when a paper is widely known and has obtained many citations the importance of the other mechanisms will increase (authors citing authoritative papers, the bandwagon effect, etc.). In other words, visibility dynamics become more important over with time because of the self-intensifying mechanisms that are involved. This explains why the relative differences in citation rates between poorly cited and highly cited papers increase over time.[5]

Another temporal effect is the phenomenon termed “obliteration by incorporation”, meaning that basic theoretical knowledge is not cited anymore. As a consequence, the most basic and important findings may not be among the most highly cited papers because they have been rapidly incorporated into the common body of accepted knowledge. One example is the paper by Watson and Crick on the double helix structure of DNA (1953). In contrast to the Lowry paper cited more than 250,000 times this paper has ”only” been cited

[5] This might also be considered as an argument for not selecting too long citation windows when constructing citation indicators (although there are other arguments for not selecting a window which is too short).

1,400 times through 1990 (Garfield & Welljams-Dorof, 1992). In the latter case the findings were rapidly incorporated into the common body of accepted knowledge and researchers would not need to explicitly cite the paper anymore. In the case of Lowry, the paper would also represent a part of the accepted body of knowledge. However, in this case as a consequence of referencing norms, it has been cited (almost) every time the method has been used. According to Lowry himself: "It just happened to be a trifle better or easier or more sensitive than other methods, and of course nearly everyone measures proteins these days" (quoted in Garfield (1979b).

At the level of individual papers one obviously finds large variation in the relative importance of quality versus visibility dynamics (if one can apply these dynamics to individual papers at all). Visibility dynamics works to increase the visibility of some papers, and to decrease the visibility of poorly cited papers. Thus, citation patterns exaggerate the differences between the bottom and the top. For any particular paper, a great many contingencies occur and will determine what happens (see Article VII for examples).

Although one can only speculate on their overall relative importance, I suggest that the relative weight of quality and visibility dynamics will be of the order of 2:1. In other words, citation indicators – in the aggregate – will reflect quality, but one-third will be due to visibility dynamics which one might want to correct for. It is reasonable to attribute the most weight to quality dynamics because, after all, a positive correspondence was identified between citation counts and research quality aspects in many studies. Thus, to a reasonable extent the citation patterns reflect the cognitive hierarchy among scientific contributions under the premise that relative, field-specific citation standards are being used. This claim does not hold, however, for individual papers. Even in the aggregate, the effects of visibility dynamics are not insignificant compared to those of quality dynamics. They are definitely not essentially random influences which average out at aggregated levels (Martin & Irvine, 1983). This is the structural reason why highly cited papers and citations more generally, have important limitations as measures of scientific performance.

1.1.3 The interpretation of citation indicators, limitations and problems

As we have seen, studies of the citation process have not provided any simple answer to the question of what citations stand for. In fact, certain characteristics of scientists' referencing behaviour have been used to question the use of citations as indicators, in particular by MacRoberts and MacRoberts (see MacRoberts, 1997; MacRoberts & MacRoberts, 1987; 1989; 1996). Based on empirical studies they have argued that only a relatively small proportion of a scientist's influences are actually cited. Moreover, the citing is biased: some sources are cited essentially every time they are used, while other research is never cited even though it may be used more than the highly cited work. In their view this means that the basis for using citations as performance indicators is undermined, and they criticize citation analysts who: "in spite of an overwhelming body of evidence to the contrary […] continue to accept the traditional view of science as a privileged enterprise free of cultural bias and self-interest and accordingly continue to treat citations as if they were culture free measures" (MacRoberts & MacRoberts, 1996).

The MacRoberts' views have led to much debate, and rightly so, but their conclusions are seen as too sweeping. Garfield, for example, has argued that it would be impossible to cite all former literature on a particular topic. The fact that authors do not cite all their influences does not invalidate the use of citations as performance measures when enough literature is taken into account (see Garfield, 1997b). While most citation analysts appear to agree that citing or referencing is biased, the question is which implications this has. For example, it has been argued that it is not fatal for the use of citation as indicators and that to a certain extent the biases are averaged out at aggregated levels. The existence of various cognitive meanings of citations and motivations for citing does not necessarily invalidate the use of citations as (imperfect) performance measures (Luukkonen, 1990; Zuckerman, 1987). Motives and consequences are analytically distinct.

A different approach to the question is represented by studies analysing how citations correlate with peer reviews. In these studies judgements by peers have been typically regarded as a kind of standard by which citation indicators can be validated. The idea is that one should find a correlation if citations legitimately can be used as indicators of scientific performance (which assumes that peer assessment can indeed identify quality and performance without bias – a dubious assumption). Generally, most of the studies seem to have found an overall positive correspondence although the correlations identified have been far from perfect and have varied among the studies (see e.g. Vinkler, 1998).

This brief comment indicates a deeper problem: what is it that a citation indicator indicates? This is not immediately given by the construction of the indicator, already because of the different dimensions of the citation process (section 1.1.2). Various concepts have been used as what citation counts might indicate, for example quality, influence, importance, significance, utility, relevance or impact. Early citation analysts such as the Cole brothers frequently referred to citations as a measure of quality, although in the introduction to their book on social stratification in science a slightly more cautious definition was given: "The number of citations is taken to represent the relative scientific significance or 'quality' of papers" (Cole & Cole, 1973).

The concept of quality is still used in the interpretation of citation indicators. Today, however, other concepts – particularly that of "impact" – are usually applied. One reason is that quality is often considered as a diffuse or at least multidimensional concept. For example, the following description is given by Martin and Irvine (1983): "'Quality' is a property of the publication and the research described in it. It describes how well the research has been done, whether it is free from obvious 'error' […] how original the conclusions are, and so on." Here, one sees reference to the craft of doing scientific research, and to the contribution that is made to the advance of science.

The impact of a publication, on the other hand, is defined as the "*actual* influence on surrounding research activities at a given time." According to Martin and Irvine it is the impact of a publication that is most closely linked to the notion of scientific progress – a paper creating a great impact represents a major contribution to knowledge at the time it is published. If these definitions are used as the basis it is also apparent that impact would be a more suitable interpretation of citations than quality. For example, a 'mistaken' paper can nonetheless have a significant impact by stimulating further research. Moreover, a paper by a recognised scientist may be more visible and therefore have more impact, earning more citations, even if its quality is no greater than those by lesser known authors (Martin, 1996).

In addition to the fundamental problems related to the multifaceted referencing behaviour of scientists, there are also more specific problems and limitations of citation indicators. Some of these are due to the way the ISI database is constructed. First of all, it is important to emphasis that only references in ISI-indexed literature count as "citations". For example, when articles are cited in non-indexed literature (e.g. a trade journal) these are not counted. This has important consequences. Research of mainly national or local interest, for example, will usually not be cited in international journals. Moreover, societal relevance, such as contributions of importance for technological or industrial development, may not be reflected by such counts. Because it is references in (mainly) international journals which are indexed, it might more appropriate to restrict the notion of impact in respect to citation indicators to impact on international or "mainstream" knowledge development.

There is also a corresponding field dimension. For example, LePair (1995) has emphasised that "In technology or practicable research bibliometrics is an insufficient means of evaluation. It may help a little, but just as often it may lead to erroneous conclusions." For similar reasons the limitations of citation indicators in the social sciences and humanities are generally more severe due to a less centralised or a different pattern of communication. For example, the role of international journals is less important and publishing in books is more common: older literature has a more dominant role and many of the research fields have a "local" orientation. In conclusion, citation analyses are considered to be most fair as an evaluation tool in the scientific fields where publishing in the international journal literature is the main mode of communication.

Then there are problems caused by more technical factors such as discrepancies between target articles and cited references (misspellings of author names, journal names, errors in the reference lists, etc.), and mistakes in the indexing process carried out by Thomson ISI (see Moed, 2002; Moed & Vriens, 1989). Such errors affect the accuracy of the citation counts to individual articles but are nevertheless usually not taken into account in bibliometric analyses (although their effect to some extent might "average out" at aggregated levels).

In addition a large number of other more specific factors have been argued to undermine the use of citations as performance measures (see e.g. Seglen, 1997). Some of these relate to the citation process, for example so-called "negative" citations, "citation circles" (groups of researchers who cite one another's work), self-citations, others to the construction of the ISI-database, for example an English language bias of the ISI database.

While some of the problems are of a fundamental nature, inherent in any use of citations as indicators, other may be handled by the construction of more advanced indicators. In particular, because of the large differences in the citation patterns between different scientific disciplines and subfields, it has long been argued by bibliometricians that relative indicators and not absolute citation counts should be used in cross-field comparisons (Schubert & Braun, 1986; Schubert & Braun, 1996; Schubert, Glänzel, & Braun, 1988; Vinkler, 1986). For example, it was early emphasised by Garfield that: "Instead of directly comparing the citation counts of, say, a mathematician against that of a biochemist, both should be ranked with their peers, and the comparison should be made between rankings" (Garfield, 1979a: 249). Moed *et al.* (1985) similarly stressed that: "if one performs an impact evaluation of publications from various fields by comparing the citation counts to these publications, differences between the citation counts can not be merely interpreted in terms of (differences between) impact, since the citation counts are partly determined by certain field-dependent citation characteristics that can vary from one field to another". Various reference standards and normalisation procedures have been developed. The most common is the average citation rates of the journal or field in which the particular papers have been published. An example of an indicator based on the journal as a reference standard is the *Relative Citation Rate*. Here the citation count of each paper is matched to the mean citation rate per publication of the particular journals (Schubert & Braun, 1986). This means that the journals are considered as the fundamental unit of assessment. If two papers published in the same journal receive a different number of citations, it is assumed that this reflects differences in their inherent qualities (Schubert & Braun, 1993).

Another dimension to the issue is the level of analysis, ranging from the individual papers to entire nations and research fields. The problems and limitations of citation analysis arise differently at these different levels of aggregation. The research group might be considered as the functional unit of science (at least within the natural sciences and medicine) and therefore particularly appropriate for an evaluation or a citation analysis. However, the research group usually represents a relatively low level of aggregation in terms of numbers of papers. For example, self-citations then represent a potentially larger problem than at higher level of aggregations (cf. Article III). Studies of research groups and departments are usually delineated in terms of publications found by searching for persons or addresses. However, author addresses are usually not standardised and the occurrence of homonymies may cause severe problems when using author names as basis for identification of publications. In consequence, a very considerable effort is usually required to identify and clean up the data before one can produce reliable publication and citation indicators for research groups and departments. When the unit being analysed is small the indicators are also more sensitive to missing publications (particularly critical when these are highly cited).

At higher levels of aggregation such as fields and nations, the situation is different. Some of the problems at micro level will not appear at all, or average out at aggregated levels. The indicators are more robust towards missing publications and self-citations are for example less likely to cause systematic bias. It is usually assumed that the validity and reliability of citations as performance indicators increases with the number of publications being analysed. Nevertheless, there are problems such as field delimitation that are particularly relevant here, for example when assessing national performance in a research subfield (cf. Article IV).

Although some critics such as the MacRoberts have rejected any use of citations as indicators because of the problems described above, the view generally held among bibliometricians is that citations represent a good but not perfect measure of research performance. This view may be exemplified by the following statement by Welljams-Dorof: “In general, the larger the citation data set being used, the higher the confidence level of the results. Analyses involving entire fields of research, nations, regions and large universities are virtually unaffected by the concerns and caveats about citation data […]. The confidence level at these large aggregate levels is quite high in analyses of fundamental, basic research. But in engineering, technology and other applied sciences, the confidence level is moderate. In these applied areas, publishing research in peer-reviewed journals is not as strong a professional tradition or motivation as in the basic sciences. Nevertheless, citation analysis still provides valuable perspectives on major contributions in the applied sciences” (Welljams-Dorof, 1997).

1.2 The use of citations as indicators in science policy and research assessments

Originally, bibliometric studies were pursued for two main reasons: for library and information purposes, and as historical and sociological studies of science. Prior to the mid-1970s, the potential of bibliometrics for science policy purposes was only seen by a few analysts (Martin, 1996). Later, science policy became an important area of application of bibliometric analysis. In fact, science policy has been very influential on the development of bibliometrics as a field of study. Basically, in science policy contexts bibliometric indicators have been used for two main purposes: for evaluating research and for monitoring research systems. While the number of citations usually is considered as an indicator of scientific impact, the number of publications is regarded as a quantitative measure of the research output.

The use of indicators in science policy is related to a general trend which demands for greater accountability of science. In this perspective evaluations and performance indicators are seen as ways in which to assure the government and the public that public monies are being well spent. One reason why evaluations have become increasingly necessary during recent years is that many countries are witnessing increasing constraints on public expenditure, including spending on research (Martin, 1996). Because of this it is becoming more difficult to find the funds to support new research areas and new scientists as well as to pay for more sophisticated and expensive scientific equipment. Moreover, the essential elite character of science, in which researchers contribute in a varying degree to scientific progress, makes it important from a science policy perspective to identify and promote high quality research. Against this background citations and other bibliometric indicators have been seen as useful.

Below some examples are given of usage of bibliometric indicators in a science policy context.

1.2.1 Usage of bibliometric indicators – some examples

<u>Monitoring scientific development – science and technology indicators reports</u>

Citations and other bibliometric indicators are often used for monitoring scientific developments. Such use may include analyses of trends in the publication activities for particular scientific disciplines and countries, for example the level of the activity in terms of number of publications and the citation rates of these publications. Typically, comparisons are made between countries and rankings. Information of this type cannot generally be provided by panels of peers since their expertise will not stretch further than a qualitative view (Van Raan, 1993). In this way such analyses may give new insights into the structure and development of national research systems.

One conspicuous example of the use of citations and other bibliometric indicators in science policy contexts is the so called "science and technology indicators reports". Today a large number of countries, as well as the European Commission, publish such reports on a regular basis. Also Norway has published such reports.

The first indicator report was issued by the National Science Board in the U.S. in 1972 (Cozzens, 1997).[6] As stated in the introduction of this first report, it aimed at developing "a system of indicators for describing the state of the entire scientific endeavour. These indicators […] are intended to measure and monitor U.S. science – to identify strengths and weaknesses of the enterprise and to chart its changing state" (quoted

[6] The publication of the report also led to reflection on science indicators by various scholars, resulting in a book published a few years later (Elkana *et al.*, 1978).

in Narin, Hamilton, & Olivastro, 2000). Later, more than a dozen such reports have been published in the US. The latest edition (2002) had more than 1,000 pages.

The indicator reports contain a broad range of input and output indicators, for example on education, human resources, funding, technology and innovation. Bibliometric data are used to construct indicators of the output and impact of the research system. Examples of indicators include number of publications, citation measures, and collaboration indicators based on co-authorship. Indicators are constructed at an overall as well as discipline level, and based on absolute and relative measures. Usually, the analyses have the main focus on international comparisons and the ranking of countries.

Because of their wide distribution, it is fair to say that through these reports bibliometric indicators play an important role in the public perception of the scientific performance of a country. According to Wouters (1999), the regular production of these highly visible reports in the Netherlands has encouraged the use of bibliometric indicators, also in lower-level evaluations. It should be added that bibliometric macro-analyses are also presented in various other publications such as policy documents, reports and specialised and multidisciplinary journals (e.g. May, 1997).

Use in research evaluations

Citations and other bibliometric indicators have been applied in various ways in research evaluation. For example, such indicators are used to provide information on the performance of research groups, departments, institutions or fields. According to van Raan (2000), "the application of citation analysis to the work – the *oeuvre* – of a group as a whole over a longer period of time, does yield in many situations a strong indicator of scientific performance, and, in particular, of scientific quality". As a qualifying premise it is emphasised, however, that the citation analysis should adopt an advanced, technically highly developed bibliometric method. In this view, a high citation index means that the assessed unit can be considered as a scientifically strong organisation with a high probability of producing very good to excellent research.

The Netherlands has been a leading country in the development and application of bibliometric indicators in the context of research evaluation. Since 1992 university research in the Netherlands has been evaluated on a regular basis (Van der Meulen, 1997). Each discipline has been evaluated every five years. Use of bibliometric indicators has been a common practice in these evaluations. Some bibliometric analyses were carried out by The Centre for Science and Technology Studies (CWTS) at Leiden University (see e.g. VSNU, 1996). The bibliometric analysis is carried out at a university and research group level. A varied set of indicators can be included, but it is up to the VSNU disciplinary panels to offer guidelines. As examples van Raan (1996) lists the following:

- Number of publications
- Number of citations received
- Citations per publication (average)
- Citations per publication, self-citations not included
- Percentage of papers not cited during the time period considered
- Average citation rate of journal set
- Average citation rate of (sub)field(s)
- Citations per publication, compared to journal set world average
- Citations per publication, compared to (sub)field world average
- Ratio of journal set world average and (sub)field world average
- Percentage of self-citations

It should be noted that this represents an advanced set of indicators compared to the simple count of raw citations that used to be, and sometimes still is, presented as a performance measure particularly by non-experts.

According to van der Meulen (1997), there is now hardly any dispute about the added value of these data although this was not the case in the 1980s. For example, in one of the evaluation reports the panel made the following statement: “the Committee considers the bibliometric data and analysis […] to be a valuable tool that complements the assessment procedures based on written and oral information. The Committee feels that the combination of the two methods has increased the reliability of the assessments” (VSNU, 1996).

A fundamental limitation of citation indicators in the context of research assessments is, however, that a certain time window is necessary for such indicators to be reliable, particularly when considering smaller number of publications. Frequently, a three-year period is considered as appropriate (see e.g. Moed, Burger, Frankfurt *et al.*, 1985). But for the purpose of long-term assessments a longer period is required. At the same time, an excessively long period makes the results less usable for evaluation purposes. This is because one then only has citation data for articles published many years previously. Citation indicators are not very useful when it comes to publications published very recently, a principal limitation of such indicators being that they cannot provide an indication of present or future performance except indirectly: past performance correlates with future performance (Luukkonen, 1997a). It should be added, however, that this time limitation does not apply to the bibliometric indicators based on publication counts.

Journal impact factors

The journal impact factor is probably the most widely used and well-known bibliometric product. It was originally introduced by Eugene Garfield as a measure of the frequency with which the average article in a journal has been cited. In turn, the impact factor is often considered as an indicator of the significance and prestige of a journal (Glänzel & Moed, 2002).

The examples of uses are numerous. The product is applied by libraries for journal selection (e.g. which journal to subscribe to), in bibliometric studies, by publishers to promote their journals, in research evaluations and by faculty selection committees as part of the evaluation of individual performance (Garfield, 1998a). Moreover, the impact factors are sometimes applied by scientists as ranking lists of where to submit their publications (Garfield, 1997a).

The journal impact factor is, however, among the most debated bibliometric indicators. This is not the place to discuss this issue in any detail (for a review see Glänzel & Moed, 2002). Some points of criticism can nevertheless be mentioned. First, the time period used as basis for the calculation of impact factor is often considered to be too short. In the standard product the impact factor is calculated as the mean number of citations in a given year, to journal items published during the preceding two years. Thus, fast-moving fields, that is, fields in which the research front is rapidly developing will be in favour. Second, there is no normalisation of differences in citation rates between fields. This means that journals in fields with high average citation rates tend to dominate the top of the ranking lists. Third, various more technical issues concerning the basis for calculating such measures have been criticised, for example on what counts as a citable item. Finally, severe criticism has been directed towards using journal impact factor for evaluating research, in particular to use these rates as substitutes for missing citation data of the publications.

The apparent simplicity, comprehensibility and availability of the journal impact factor explains why it has achieved this broad popularity. However, considering the serious methodological shortcomings and problems with using its popularity may seem rather unfortunate.

Examples from Norway

As the studies of the dissertation have been based on Norwegian data, some examples of use of bibliometric indicators in Norway are appropriate. I have also selected a few cases in which the problems and limitations of such indicators are illustrated.

Bibliometric studies have been carried out in Norway since 1985, mainly by the Norwegian Institute for Studies in Research and Higher Education (NIFU STEP). Studies undertaken have varied from basic research projects to the production of simple publication counts included in various reports. Because of the institute's science policy orientation, most of the work has been related to science policy issues either directly or indirectly. Bibliometric indicators have been used both to monitor scientific development and as part of research evaluations. An example of the first type of application is the bibliometric indicators included in the biannual Norwegian science and technology indicator reports. In the research evaluations, bibliometric analyses have usually only been carried out at an overall national level. Thus, such indicators have not been used in the assessment of the research performance of individual research groups and departments. Nevertheless, there have been a few cases where the panels have attempted to carry out their own bibliometric analyses.

One example of this is an evaluation of a medical research institute in Norway (see Sivertsen, 1997). Here, without assistance of bibliometric expertise, the evaluation committee applied a citation indicator which ranked individual scientists according to the impact factor of the journals in which they had published. It appeared that the ranking method was highly problematic. In fact, one of the professors evaluated became a bibliometric expert himself in reaction to this evaluation (Seglen, 1989). By counting the actual citations of all publications involved in the evaluation and doing a statistical analysis, he was able to demonstrate that the indicator was highly misleading. On his publication list this professor had a so-called citation classic – an article that had received more than a thousand citations. About six citations were allocated to this article when the journal impact factor was used.

There are also other cases of problematic applications, showing the importance of understanding the pitfalls the limitations of such indicators. For example, some years ago

Scandinavian newspapers published a trend line showing a significant decline in the number of articles of Scandinavian chemistry during the previous fourteen years (Sivertsen, 1997). The data was based on ISI data processed by Computer Horizons, Inc. and was interpreted by the journalists as a decline in the productivity of Scandinavian chemistry. However, it was subsequently revealed (Sivertsen, 1992) that the decline could largely be ascribed to a reorganisation of Scandinavian chemistry journals, particularly to an abandonment of the domestic journals. Thus, the data could not reasonably be interpreted as a decline in productivity.

Another example concerns the calculation of overall citation rates for countries which is one of the most popular bibliometric indicators. It is used in a variety of contexts and is included in standard bibliometric products such as ISI's *National Science Indicators* (NSI). Basically the indicator is expressed as the inverse relationship between the total number of papers published by a country and the number of citations received by those papers within a chosen time period. However, the scientific profile of countries varies considerably. The large differences in citation rates between scientific fields means that countries specialising in highly cited field will have a comparative advantage. Because the indicator is highly influenced by the countries' relative scientific specialisation (which in any case cannot be considered to be related to their scientific performance) NIFU STEP has developed an alternative method of calculation. This method adjusts for the differing field composition among countries and is published in reports such as the Norwegian science and technology indicator report. Accordingly, it gives a somewhat differing picture than the "standard" indicator.

In Sweden the overall citation rate of the country has showed a declining trend for several years (Persson, 2000). This has caused much concern. The indicator used, has been the "standard" indicator of the NSI database described above. In the scientific newspaper *Dagens Forskning* (2003) a professor wrote an article arguing that this impression was largely wrong, referring to a figure in the Norwegian science and technology indicator report showing a more positive trend for Swedish science. Apparently, a significant share of the Swedish decline could be ascribed to a changing field structure in which Sweden adapted to a world average field structure (a reduced relative activity in the highly-cited fields). This example may be taken as another illustration of the importance of understanding the limitations of the database and the way the indicators are calculated when the indicators are being interpreted (compare the similar debate on the decline of

British science: Braun, Glänzel, & Schubert, 1989; Irvine, Martin, Peacock, & Turner, 1985).

1.2.2 Bibliometric indicators, peer review and the science system

Today most bibliometricians emphasise that a bibliometric analysis can never function as a substitute for a peer review. Thus, a bibliometric analysis should not replace an evaluation carried out by peers. This is due to the many problems and biases attached to such analyses. As a general principle, it has been argued that the greater the variety of measures and qualitative processes used to evaluate research, the greater is the likelihood that a composite measure offers a reliable understanding of the knowledge produced (Martin, 1996).

At the same time, it is generally recognised that peer reviews also have various limitations and shortcomings (Chubin & Hackett, 1990). For example, van Raan (2000) argues that subjectivity is a major problem of peer reviews: The opinions of experts may be influenced by subjective elements, narrow mindedness and limited cognitive horizons. An argument for the use of citation indicators and other bibliometric indicators is that they can counteract shortcomings and mistakes in the peers' judgements. That is, they may contribute to fairness of research evaluations by representing "objective" and impartial information to judgements by peers, which would otherwise depend more on the personal views and experiences of the scientists appointed as referees (Sivertsen, 1997). Moreover, peer assessments alone do not provide sufficient information on important aspects of research productivity and the impact of the research activities (Van Raan, 1993).

In this way a bibliometric study is usually considered as complementary to a peer evaluation. Van Raan has accordingly suggested that in cases where there is significant deviation between the peers' qualitative assessments and the bibliometric performance measures, the panel should investigate the reasons for these discrepancies. They might then find that their own judgements have been mistaken or that the bibliometric indicators did not reflect the unit's performance (Van Raan, 1996). [7]

[7] Van Raan (1996) suggests that in cases were conflicting results appear, the conclusion may depend on the type of discrepancy. If the bibliometric indicators show a poor performance but the peer's judgement is positive, then the communication practices of the group involved may be such that bibliometric assessments do not work well. By contrast, if the bibliometric indicators show a good performance and the peers' judgement is negative, then it is more likely that the peers are wrong.

Nevertheless, although citations may be valuable in such contexts, citation indicators have sometimes been used simplistically by policy makers. While, on the one hand, there has been a development towards more advanced and sophisticated indicators (e.g. Van Raan, 1996) there are also many examples of uninformed use, of tendentious use (i.e. selecting the most "advantageous") and misuse of citation indicators. This happens increasingly because the use of such data has spread beyond the bibliometric community, and where examples of use include those by tenure and promotion committees, funding and policy agencies and journalists (Welljams-Dorof, 1997). A survey by Hargens and Schuman (1990), for example, revealed that a substantial proportion (35%) of biochemistry departments and the majority (60%) of sociology departments at US universities used citation counts when making decisions about appointing, promotion and tenure.[8] As a response to an increasing demand for bibliometric indicators Thomson ISI has started to produce their own evaluation tools (e.g. *ISIHighlyCited.com*) that are accessible online to anyone with an internet access to a university library. As a consequence it is also possible for scientists and research administrators lacking bibliometric competence to produce seemingly objective assessments of researchers and their publications (see Weingart, 2005).

Even when professional bibliometricians produce reports and analyses these are not (despite the recommendations) normally combined with a corresponding peer assessment. Nevertheless, based on the indicators, conclusions are drawn on various aspects of scientific performance. Recently, ranking of universities based on simple bibliometric measures has attracted much attention (Van Raan, 2005). Although a certain disparity between the recommendations and the common practice may exist, the recommendations would be primarily directed towards micro-level applications or evaluations in which the results have direct consequences for the units assessed (e.g. in terms of funding). Anyhow, because publication and citation indicators have been and will continue to be used alone it is even more important to be aware of the limitations of such indicators.

Further, one often encounters opposition towards the use of bibliometric indicators for evaluation purposes in many scientific communities. In particular, if the evaluations have consequences for research funding, scientists are concerned about their possible lack

[8] The higher percentage for the sociology departments may be seen as deriving from the larger uncertainty about quality and performance in at least some of the social sciences as compared with the life sciences, and thus a larger need for objectifying measures.

of fairness. If evaluations are critical or negative they are likely to generate protests, although this applies to all evaluations regardless of methods used (Luukkonen, 1997a).

The use of citations as indicators in research evaluations has also had an impact on the scientific system itself. For example, publishing in journals with a high impact factor or in journals indexed by ISI has become an independent measure of scientific quality (Wouters, 1999).

Clearly, the increasing use of indicators may lead to changes in the behaviour of scientists and thus in the characteristics of the system being measured. Some of these may be intended changes, for example when a research institute uses funding formulas as a mean to stimulate more publishing in international journals. But there are also unintended and unwanted consequences because scientists adapt to the measurement system, for example by slicing papers into the "smallest publishable unit", illegitimate use of co-authorships, and so forth. Because of this it has been suggested by Martin (1992) that there is a form of Heisenberg Principle at work here – if one attempts to measure a research system it becomes disturbed and the measurement will reflect this. For example, it was observed in Australia that the use of formula based on publications (and other measures) lead to a significant increase in the publication output of Australian academics. However, because quality was paid scant regard in the measures, there was little incentive to strive for the top journals. Accordingly, the biggest increase of publications was in the journals at the lower end of the impact scale (Butler, 2003). It has been argued that when many different indicators are being applied it is much more difficult, if not impossible, to manipulate all the indicators without at the same time improving one's own research (Martin, 1996).

Some authors have been very critical of the impact of bibliometrics upon the science system (Weingart, 2005). A special concern is the situation in which citation counting alone becomes the dominant method of evaluation (Luukkonen, 1990) because in such a situation important research benefits to society (technological spin-offs, economical benefit, manpower training, etc.) are neglected.

In sum, the use of citations as performance measures have their limitations, as all bibliometric indicators have. But a citation analysis when well designed and well interpreted will still provide valuable information in the context of research evaluation. Performance, quality and excellence can also be assessed through peer review, but in spite of their widespread use, these have problems as well. A combination of methods, or better, mutual interrogation on the basis of findings of each of the methods, is more likely to provide reliable evaluation results.

Already in the 1980s, analysts emphasised that there is no straightforward nor simple relationship between the results of bibliometric analyses and science policy decisions (e.g. Moed *et al.*, 1985). However important it is to further improve citation and other bibliometric indicators for use in evaluation and policy, just as important is insight into the factors and mechanisms underlying the observed bibliometric patterns. Section 1.1.2 has shown this for the citation process. This thesis collects a number of studies which analyse bibliometric patterns in such a way that factors and mechanisms can be traced.

2. Highly cited articles and a preview of the studies in the subsequent chapters

The previous chapter outlined the general background associated with citations and bibliometric analysis from which the thesis arises. In this chapter, the central issues of skewed citation distributions and highly cited papers (and the use of citations as indicators) are discussed more specifically so as to provide a starting point for an overview of the subsequent chapters (2.2) and a brief discussion of data and methodology (2.3). Based on different projects the thesis explores an intersect between the bibliometric and sociological questions about the phenomenon of citations, and the science policy issues about using such data as indicators in decision-making.

2.1 Highly cited papers and citation indicators

It is well-known that there are large differences in productivity between scientists: a relatively small proportion of scientists contribute to the majority of the publications. In 1926 Lotka formulated his famous inverse square law of productivity, which states that the number of authors producing *n* papers is approximately $1/n^2$ of those producing one (Lotka, 1926). This means, for example, that of all authors in a given field, 60 per cent will have produced just one publication. The results of several later studies have, however, shown that productivity differences in scientific publishing are less than indicated by Lotka, and that Lotka’s law overestimates the number of papers produced by the most prolific scientists. Nevertheless, there exists a highly skewed pattern of productivity in scientific publishing (Kyvik, 1991: 90).

As noted already in Chapter 1 citation distributions are also extremely skewed (see Garfield, 1990; Seglen, 1992). This is clearly exemplified in the scientific production of Norway: among 50,000 Norwegian ISI-indexed articles published in the period 1981–1996, almost one half have never been cited or have only been cited once or twice, while 1% of the papers have been cited more than 50 times within five years after publication. Similar distribution patterns can be found for all countries. The skewness characterising these bibliometric distributions are striking and have also been the origin of much discussion, particularly concerning the science policy implications of skewness.

In 1972 Jonathan and Stephen Cole published a paper critisising the so-called "Ortega Hypothesis" (Cole & Cole, 1972). This hypothesis states that minor contributions made by "average" scientists are the important and necessary basis from which major scientific advances and breakthroughs arise. The Cole brothers argued that most of the works published have little or no impact on the development of scientific knowledge. Relying on studies of citation and publication patterns within physics showing skewed distributions, they claimed that the driving force behind scientific progress is powered by relatively few articles in a given field. In other words, only a small proportion of the scientists actually contribute to scientific advance through their published research (see also Cole, 1970).[9] In addition, some very controversial policy implications were considered, in particular that it would be possible to significantly reduce the size of science without slowing down the rate of scientific advance. [10]

Particular attention has been given the phenomenon 'uncitedness'. In 1990 an article in *Science* reported that 55% of scientific articles did not receive a single citation within five years of publication (Hamilton, 1990). In addition, large variations in the rate of uncitedness between disciplines were identified, rising as high as 98% in the humanities (Hamilton, 1991). This was striking, although later studies showed that limiting the sample to articles only (eliminating "marginalia" such as book-reviews, editorials, letters, etc.) and excluding the social sciences and the humanities, would reduce uncitedness to 22% (Schwartz, 1997). In the debate that followed it was (again) argued on the basis of the figures that half the scientific work is basically worthless, calling for more rigorous standards in funding and publishing practices.

Using citation patterns for such conclusions, however, is clearly controversial (Chubin & Hackett, 1990). When the phenomenon of large-scale uncitedness could be traced empirically, it had to be interpreted. Some continued to stress the Ortega hypothesis that small incremental additions to knowledge embodied in papers receiving few or no citations is also essential to scientific progress. MacRoberts and MacRoberts added intentional neglect when they argued that many people contribute significantly to scientific

[9] In later works S. Cole has concluded that the number of scientists who will make significant contributions to science is in general a linear function of the total number of people entering the field of science (Cole, 1992: 225).

[10] One counter-argument was that one does not know which parts of science to stop. Another counter-argument took the research rather than the citedness as starting point: if half of the publications are not cited, this means waste of good research (Turner & Chubin, 1976). Their later paper adds a discussion about chance in science (Turner & Chubin, 1979).

progress but receive little or no credit by way of citations for their contributions (MacRoberts & MacRoberts, 1987). In other words, scientists must frequently fail to make explicit reference to much of the work that has influenced them.

During the last decade there has been an emerging interest towards using the top end, the highly cited papers, as indicators in research assessment. One reason for this is the increasing focus on scientific excellence in science policy (Van Raan, 2000). In many countries, including Norway, this is exemplified by the initiative for establishing centres of excellence. In this context, highly cited papers have been regarded as potential candidates for identifying and monitoring “excellent” scientific research. Lately, this was shown in a benchmarking study from the European Commission in which highly cited papers were used as indicators for comparing the research performance of the EU countries (European Commission, 2001). As other examples, Thomson ISI regularly produces data on highly cited papers and researchers (*ISIHighlyCited.com*), while in the journal *Current Contents* the authors of so-called “citation classics” have been asked to comment on their highly cited papers. Highly cited papers have also been applied as indicators in case studies of research groups (e.g. Martin & Irvine, 1983) and an explorative study by Tijssen *et al.* (2002) concluded that highly cited research papers represent useful indicators for identifying “world-class” research. Garfield found that high rankings by citation frequencies were positively correlated with Nobel prizes in the way that nearly all Nobel laureates were highly cited within their disciplines and had produced highly cited papers (Abt, 2000; Garfield & Welljams-Dorof, 1992). Moreover, various studies and research on highly cited papers have also been carried out (Campanario, 1993; Campanario, 1996; Cano & Lind, 1994; Glänzel & Czerwon, 1992; Glänzel, Rinia, & Brocken, 1995; Glänzel & Schubert, 1992; Oppenheim & Renn, 1978; Plomp, 1994).

In the view of Small it can be assumed that frequently cited papers represent the key concepts, methods, or experiments in a field. Highly cited papers have been viewed as ”exemplars” (using Thomas Kuhn’s terminology), whereby papers are cited because they represent a classical study, a ”concept” marker (Small, 1978), or show how a particular line of research is carried out. This was also a conclusion in a study of highly cited papers in psychology (Shadish, Tolliver, Gray, & Gupta, 1995). Here, the characteristics of an exemplar were that its author was a recognized authority; and that the work was thought to be classic, early reference that represented a genre of studies which generated novel research and that resisted falsification.

Nevertheless, at the level of the individual paper, frequent citations do not necessarily equate with breakthrough science. But at aggregated levels such indicators have more promise with respect to assessing aspects related to scientific excellence (Tijssen *et al*., 2002). The concept of excellence is complex however (Tijssen, 2003), and a further discussion of the relation between highly cited papers and scientific excellence would require a more precise definition of what is meant by excellence.

One possibility is to think of contributions to scientific advancement. This then raises a further question, at first sight of a technical nature. The number of citations to important or useful papers in poorly cited fields will tend to be much lower than to similar papers in highly cited fields. The solution of using relative standards in identifying highly cited papers seems obvious, but does imply a decision to consider both types of contributions as equal. There is a corresponding issue related to the field size. Important or useful papers in small, narrow or highly specialised fields will tend to receive far fewer citations than similar papers in larger fields, or compared to papers addressing more general issues of relevance within many fields. Using reference standards based on field averages will not eliminate such differences because the average citation rate is not influenced by the size of the field. In other words, for being highly cited it is a disadvantage to work on highly specialised issues. Again, one might argue that one is justified in considering general contributions as more important than more specialised contributions. On the other hand, scientific progress depends on both kinds of contribution and it is not automatic that the first kind should be given more “value” than the latter. At least one should reflect on such possible implications before using highly cited papers as indicators.

2.2 The empirical studies: a preview

Against this background this dissertation makes further contributions to the issues of skewed distribution, highly cited papers, and citation indicators. One issue is what one is “measuring” using highly cited papers and how these measures relate to various aspects or conceptualisations of scientific excellence. Another issue is that when citations are used as indicators, the effect of highly cited papers is usually not given attention – notwithstanding the fact that due to the skewness of citation distributions highly cited papers, may have large effect on the outcome.[11]

In one of the studies in this dissertation (Article I) the latter issue is given particular attention. Here, the underlying citation distributions of national citation indicators are further explored. The main purpose is to analyse how the citation mean is affected by a few highly cited papers. This study investigates and documents the skewness of citation distributions and discusses its effect on indicators.

A second study in this dissertation (Article II) analyses the bibliometrical characteristics of highly cited papers. This has relevance for how the citation indicator should be interpreted and how highly cited papers may be used as indicators for assessing research performance. This study also discusses the conceptual distinction between quality and visibility dynamics in order to explain the skewness of citation distributions. As outlined already in 1.1.2, quality dynamics refers to the way citations capture the substance of contributions, some becoming major scientific advances; others serving to fill in the details. Visibility dynamics, on the other hand, refers to certain mechanisms that draw on incipient citation and reputation, such as the "bandwagon" and "Matthew" effects.

When citations are used as indicators various issues arise, independent from the skewed distributions of citations. Some concern more technical aspects involved in the construction of indicators; others focus on sociological issues. The citation behaviour of scientists is always a key factor in the value of citations used as indicators. One issue is self-citation, i.e. citations in which authors cite their own publications. This can be seen as undermining the use of citation counts as performance measure. In the next study of the dissertation (Article III) I focus on how prevalent self-citations are and discuss the implications of the findings. Rather than using a micro-sociological approach, here I investigate self-citation bibliometrically on a large scale basis.

The characteristics of the ISI-database affect the question of the usefulness and validity of citations as indicators. As described above some of the problems and limitations of citation indicators are due to the way the ISI-database is constructed. As issues for a further analysis I have chosen to focus on two important questions: database coverage and

[11] This issue was exemplified in a bibliometric analysis of the University of Bergen. Here, one of the nine tenured professors at a biology department collected 86 % of the total citations received by the department (over a five-year period) (Taxt & Aksnes, 2003). Using indicators based on average citation rates this department obtained a very high score. However, this was due to the production of this professor – the other publications of the department were rather poorly cited.

field delineation. Norwegian microbiology has been selected as a case for this study (Article IV).

As a way of validation, citation indicators have been compared with various types of peer reviews. The question of validity can, however, be approached from different angles. When citations are used as indicators, it is almost by definition an aggregate of papers that are being analysed, representing a university department or a scientific field, for example. One question is how citations work at such aggregated levels. This can be addressed by comparing the average citation scores with other performance measures such as peer ratings. In the next study of the dissertation (Article V) I have used such an approach, i.e. to examine the correspondence between citations indicators and assessments based on peer reviews. This study is based on an analysis of research groups at the University of Bergen, Norway.

In further study of external validation of citation counts I analyse the correspondence between citation counts of publications and the authors' own assessments of their scientific contribution (Article VI). The focus of this study is the individual publications and their citation counts. The study is based on a questionnaire survey carried out among Norwegian scientists.

Finally, Article VII addresses how citations are perceived among scientists. Based on the same questionnaire survey scientists' views on being cited and how this relates to actual contributions were investigated. This allows me to tap their experience, and their "theories" as to the fairness of citing patterns and the factors which influence and undermine citations as measures of scientific contribution.

As a general point, I note that all the studies relate to Norwegian science. In the bibliometric studies Norwegian publications, or more precisely articles in the ISI database with at least one author address in Norway, have been analysed. None of the studies include the humanities or the social sciences, because their publication and citation patterns generally differ from the other sciences. As a consequence, the questions concerning the use of citations as indicators arise differently here and would require a particular discussion.

I note also that, according to common practice, I have frequently used relative measures when addressing issues relating to the use of citations as indicators. Also, when identifying highly cited papers, I have used field-specific reference standards. The justification of this is provided by the large differences in average citation rates that can be found among fields, reflecting field-specific reference practices and technical issues relating to the coverage of the ISI-database (see Chapter 1 and 2.1).

2.2.1 What effects do highly cited papers have on national citation indicators?

This study (Article I) analyses how highly cited papers affect the citation indicator at a national level. Thus, the focus is on the use of citations to characterise a population. Basically, a national citation indicator is calculated as the average citation rate of all articles in a given field: The total number of citations in the field is divided by the total number of articles in the same field. This indicator is often used for assessing the scientific performance of a particular country, for example by comparing the results with other countries or the worldwide average. However, in such contexts the underlying citation distributions are seldom analysed. As we have seen, because of the skewness characterising citation distributions a few highly cited papers may have large effect on the citation mean. This is an issue which is important for how the citation indicator should be interpreted, but has, nevertheless, not been given sufficient attention in former studies (see however, Moed, Van Leeuwen, & Reedijk, 1999; Seglen, 1992).

Against this background this study analyses the effect of highly cited papers on national citation indicators for Norway. The analysis is based on one of ISI's standard bibliometric product, *National Science Indicator*. The effect of the most cited and the five most cited articles in various scientific fields is analysed. The basic idea is that an examination of how the average citation rate is influenced by highly cited papers may provide further knowledge on what the indicator actually indicates. Thus, an analysis of the underlying distribution may reveal other properties than one is normally aware of in the research policy discourse.

2.2.2 Characteristics of highly cited papers

The focus of this study (Article II) is the bibliometric and quantitative characteristics of highly cited papers. There have been a few earlier studies concerning this issue. Generally, highly cited papers have been found to be very different from "ordinary" cited papers. For example, it has been shown that highly cited articles tend to be authored by many researchers and are typically the result of international collaboration (Glänzel *et al.*, 1995). The purpose of the study is to examine various findings concerning the characteristics of highly cited papers, assessing their generality in respect to a sample of highly cited papers in Norwegian science. Former studies are limited in scope and numbers and new hypotheses on this topic are also being addressed. In addition, the question of how highly cited articles differ from an "average" publication are being analysed.

Examining the literature there does not seem to exist any common definition or agreement on what constitutes a highly cited paper. Frequently, a scientific publication is considered as highly cited if it has received more than a specific number of citations during a particular time period. Alternatively, scientific publications are ranked according to declining citation frequencies and a specified number or percentage are considered as highly cited (see, for example Garfield, 1990; Garfield & Welljams-Dorof, 1992). Nevertheless, the particular threshold for being highly cited has to be set arbitrarily. The large differences in citation frequencies between scientific fields means that when ranking lists are produced based number of citations, one will typically find that the top of such a list is dominated by papers from the highly cited (typically biomedical) fields. The effect is exemplified in the scientific production in Norway. Among the 100 most highly cited Norwegian articles (1981-96) we find that 92% of the publications are within life sciences/clinical medicine.

In this study relative standards are used (cf. 2.1). That is, a method of selection is applied that adjusts the field-specific differences in citation rate distributions. A publication is considered as highly cited if the number of citations received is more than a certain multiple of the mean citation rate of the particular subfield. The effects caused by field-specific characteristics will then be eliminated, and other mechanisms can more clearly be illuminated. The method of selection has similarities to the method applied by Glänzel *et al*. (1995) in a study of highly cited papers in physics. The alternative method of identifying a certain percentile of the articles (the high end) within the different scientific fields was not applied because of restriction in the available data.

2.2.3 Self-citations

This study (Article III) aims at contributing to the knowledge of self-citation, and self-citation rates in particular. A self citation (more precisely an author self-citation) is a citation in which an author cites one of his previous works. When citations are used as indicators, self-citations are often considered as problematic (MacRoberts & MacRoberts, 1989; Seglen, 1997). The reason is that self-citations do not really reflect what citations are considered to measure – the impact of a work in the scientific community. Against this background it is important to know how prevalent self-citation is and how it influences the citation indicators.

Self-citation may, however, be regarded as an ambiguous phenomenon. On the one hand citing oneself may have important functions, for example by connection the present work of an author to his/her former works within the area. Thus, given the cumulative nature of individual research, self-citation may be considered as a natural and acceptable procedure (Phelan, 1999). On the other hand, self-citation may also reflect egotism. For example, authors may tend to gratuitously cite their own works in order to raise their citation counts or to make their former works visible – although there are practical (frequency of publication) as well as normative limits for how often one can cite oneself.

A number of studies of self-citations have already been carried out (see, for example Bonzi & Snyder, 1990; Lawani, 1982; Snyder & Bonzi, 1998; Tagliacozzo, 1977). Moreover, in some bibliometric studies of research performance the percentage of self-citations has been included as an indicator. For example, a case study of physics in the Netherlands (1985-1994) found a self-citation rate of 29% (Leeuwen van, Rinia, & van Raan, 1996). Similarly, 29% of the citations in a study of Dutch chemistry (1980-1991) represented self-citations (Moed & van der Velde, 1993). However, a drawback with most former studies is their small sample size and there is need for more systematic knowledge of self-citation rates. This is particularly important for assessing how representative citation indicators are as performance measures.

In the study more than half a million citations to the scientific production of Norway (1981–1996) are analysed. The study investigates the element of self-citation in this scientific production and how the share of self-citation varies according to different parameters such as scientific discipline, overall citedness and number of authors. An additional aim of the study is to discuss the implications of the findings with respect to the use of citations as indicators.

By way of definition a self-citation is counted as a citation in which the citing and the cited paper have at least one author in common. Although it is an important advantage of the method applied that a large number of citations can be analysed, it should be added that homonymies (different authors with identical author names) cannot be identified, causing cases of erroneous self-citation counting. Contrary to this, variants or misspellings of author names, as well as change of names (e.g. due to marriage), imply that some self-citations are not recognised as such. It is not possible to identify the effects of these errors, but it seems unlikely that they would affect the overall conclusions of the paper.

2.2.4 Field delineation and the problem of database coverage

How representative are bibliometric indicators of national research activity within various scientific fields? This is a basic methodological question in bibliometrics. Since such indicators represent a major application of bibliometrics, and may function as a premise for national science policy, it is important that they build upon a valid and representative foundation. I focus on this issue in this study (Article IV).

Defining or delineating a scientific field or subfield is a general problem that not only arises in the context of bibliometrics. Even among experts there may sometimes be considerable disagreement on how a particular field should be defined, and the increasing interdisciplinary nature of scientific research makes it increasingly difficult to draw boundaries between the various fields and subfields. In bibliometrics, particular problems arise because many scientific articles deal with complex subject matters that cannot be unequivocally assigned to a single field. For instance, one and the same biomedical paper may well be justifiably classified as for example biochemistry, molecular biology, microbiology, immunology or clinical medicine. Because of this there are bound to be overlaps between subfields. Typically, when classifying scientific articles some of the papers will be easy to classify as being within the core of a particular subfield, others will be more difficult, being only weakly related to the subfield or belonging to the borderline area between different subfields.

Despite such general problems, practical working delineations are required in order to do bibliometric research. Ideally, the criteria applied should retrieve all publications belonging to a particular subfield, and at the same time be precise enough to exclude publications irrelevant to the field. Most bibliometric analyses employ journal-based subfield definitions, meaning that all articles in a given journal are assigned to the same subfield (see, for example Braun, Glänzel, & Grupp, 1995; Miquel, Ojasoo, Okubo, Paul, & Dore, 1995).

One problem with this method is that many journals contain articles dealing with a relatively broad range of fields. Consequently some articles will be included in a subfield to which they are only weakly related, whereas other, highly relevant articles will be missing. Despite the general recognition of these problems, little work has been done to examine how representative the journal-based subfield definitions actually are.

In the study I analyse how well a journal-based subject classification represents national performance within a particular scientific field. As a suitable case subfield I have chosen *Microbiology*, a typically complex biomedical discipline that includes basic as well as applied and clinical research. It corresponds to one of ISI's major subject divisions, and in this context should perhaps be regarded as a field rather than as a subfield. *Microbiology* has clear affiliations to other major ISI fields/subfields like *Biology & Biochemistry*, *Pharmacology*, *Immunology*, *Molecular Biology & Genetics*, and *Clinical Medicine*, and would, therefore, be expected to typify the problems associated with a journal-based field delineation.

A related question which also is addressed in the study of microbiology is how well the database covers the scientific production. This has long been recognised as an important methodological issue in bibliometrics. Historically, the *Science Citation Index* was constructed on the basis of Bradford's law which stated that a relatively small number of journals publish the majority of the significant scientific results. In fact, the publication data shows that just 100 journals account for 20% of what is published in ISI-indexed journals in a given year, and 40% of what is cited (Welljams-Dorof, 1997). The basic objective of ISI has been to provide a comprehensive coverage of the world's most important international journals – the core journals within the different scientific fields, but excluding the peripheral and less important journals.

Considering ISI's objective of covering the core journals, a complete coverage would not be expected. Nevertheless, it is generally known that several problems exist with respect to the coverage of the ISI database, and in particular on how well it represents the scientific production within a country or scientific subfield. A basic problem is related to the varying traditions and customs regarding publication and citation. Only in fields in which publication in international journals is the major mode of communication can such databases be expected to provide a representative picture. For example, a study of the research performance at Australian universities (Bourke & Butler, 1996) showed that the percentage of the overall scientific production represented in ISI source journals varied considerably among the disciplines. Whereas the chemical, physical, biological, agricultural and medical sciences had more than 75% of their published research output appearing in ISI journals, the corresponding percentages were significantly lower in the earth sciences, mathematics, engineering and information sciences (see also Moed, Burger, Frankfurt *et al*., 1985).[12]

The question of database coverage is not a major issue of this thesis. However, in the study of Norwegian microbiology I also examine the proportion of the scientific production that is covered by a particular ISI database. It should also be added that the study mainly focuses on publication indicators, not citations. However, the conclusions drawn are equally relevant with respect to citation indicators.

2.2.5 Peer review and citation indicators

Many previous studies have analysed how citation indicators correlate with assessments made by peers. As noted in 1.1.3, most of the studies seem to have found an overall positive correspondence although the correlations identified have been far from perfect and have varied among the studies (Vinkler, 1998). The studies also differ in methodology and levels of investigation. Luukkonen (1991), for example, found a tendency for citation counts to correlate roughly with peer ratings. In a more recent study Rinia *et al.* (1998) found that various citation indicators correlated significantly with peer ratings of research programmes in condensed matter physics. Similarly, Oppenheim (1997) identified strong positive correlations between citation indicators and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology – although his conclusions were challenged by Warner (2000). Other studies have found a correlation between citation counts and other measures of research impact or scientific recognition. For example, Cole and Cole (1973) found such a correlation in respect to Nobel prices, honorary awards and reputational ability.

As an attempt to contribute further to this issue, this study (Article V) investigates the relationship between bibliometric indicators and the outcomes of a peer review at the University of Bergen, Norway. A main question is how the peers' ratings of the scientific performance of research groups correlate with various bibliometric indicators. Citation indicators are given the main focus in the study, but also other kinds of bibliometric indicators are analysed. Particular attention is given to those cases where judgements on the basis of bibliometric results did not correspond with the views of the peers: What might

[12] In the social sciences and the humanities where books are an important publication category, the problem of database coverage is particularly great, although varying between the different subfields (see, for example Hicks, 1999; Nederhof, Zwaan, De Bruin, & Dekker, 1989; Schoepflin, 1992).

serve to explain these divergent results? The study contributes to the issue of what such indicators stand for, their limitations and usefulness as evaluation tools. In this way the study provides further knowledge on the potential and limitations of bibliometric indicators in research assessments.

As is evident the study is located within a research direction that is using judgments by peers as a kind of standard to which bibliometric results can be checked. It is important to emphasise that a number of problems also exist in respect to the fundament for such comparative studies. First of all, peer review is oriented towards a decision context: is this paper eligible for publication, is this proposal eligible for funding? Aspects related to quality might well be one of the considerations in formulating the advice, but is not the only consideration. In other words, a peer-evaluation may involve assessments of factors that are not likely to be reflected through citation counts. Only when citation indicators are used in the same decision context as peer review and the two focus on the same aspect of the scientific performance can one reasonably compare them. Secondly, judgements made by peers may not necessarily be considered as the "truth" to which bibliometric indicators should correspond – the peers may be biased or mistaken in their assessments, or they may not be competent to judge (Rip, 1997). It is therefore a question of the extent to which peer assessments and citation indicators can be compared and be expected to correlate. These issues are also addressed in the study.

2.2.6 Citation rates and perceptions of scientific contribution

While the study described above focused on indicators representing aggregated publication levels, the approach of this study (Article VI) is to look at the micro level and analyse the citation counts of individual papers. In other words such an approach implies that the building blocks of the citation indicator are examined. The basic idea is that an analysis of these building blocks will provide more knowledge about what the indicator stands for and represents.

In the study I have asked scientists about their own publications, focusing on the scientific contribution of the papers and their citation counts. One of the main purposes has been to assess how the citation counts of the publications correspond with the author's own perceptions of scientific contribution. By analysing this question for different types of contributions and different types of papers the study attempts to provide a broader picture of what citations "indicate", and the limitations and biases of citation indicators.

In other words, with the help of the respondents the study looks at cited papers "from the bottom up". One may then observe interesting things which may not be seen when looking from the outside in, as when correlation between citation rates and other measures of scientific quality and impact is studied. The advantage of asking the authors is that they have first hand knowledge of the particular publications and their research fields. On the other hand, authors' perceptions may not necessarily correspond with those of other peers. Although this is an important point that needs be taken into consideration when interpreting the results, it is also the case that other peers are fallible in their judgements (i.e. biased and mistaken). Accordingly, an "objective" yardstick with which citations can be compared does not exist, and asking the authors is one point of entry to the question.

The study has been designed as a sequel to the project on the characteristics of highly cited papers. As basis for a questionnaire survey the authors of the papers previously identified as highly cited were identified. From each of these individuals' scientific production a selection of papers to be included in a questionnaire was then made: some highly cited, some medium cited and some poorly cited. By choosing scientists that had published highly cited papers I was able to examine papers representing the full range of the citation distributions. This also enables particular attention to be drawn towards highly cited papers.

In the questionnaire the authors were asked how they assessed the scientific contribution of their papers, what types of scientific contribution the articles represented and to what extent the citation counts of the articles reflected their scientific contribution. In addition to questions on their own publications, the authors were asked some general questions concerning citations.

2.2.7 Researchers' perceptions of citations

This study (Article VII) investigates in more detail how citations are perceived by scientists. The study is the second report from the questionnaire survey described in 2.2.6 (Article VI). In this survey the scientists were also asked to give written comments on various issues such as why the citation count reflects, or does not reflect the scientific contribution of an article (generally and for specific articles) and how articles become highly cited. The scientists were also asked for comments concerning the citation counts and citation history of their own publications. This study reports the results of this part of the questionnaire survey.

Because they are not bibliometric experts the scientists' views may be regarded as expressions of "folk-theories" concerning citations. They are based on their experience of scientific publishing, communication, recognition and rewards, and the stories that are told about citations, rather than on systematic study. There may be "citation myths", and in some cases these could be identified by comparing the "folk theories" with findings from sociological and scientometric studies. On the other hand, they are the actors having real life experience of scientific publishing and communication. Thus, they may have valuable insights and knowledge concerning citations which could be the starting point for further studies. This is particularly important because of the lacunae in our understanding of citations and their role in the world of science.

The "folk theories" circle around three issues: the relation between the quality (or importance or significance) of a paper and its citation history; the importance of visibility and how all sorts of factors play a role in determining citation in general and high citation in particular; and the fairness (or lack of fairness) of the system. The paper reports and characterizes this repertoire of views and experiences about citations. Taken together, the respondents' answers and comments offer an informal (and fragmented) sociology of citations and their role.

2.2.8 The internal relation of the studies

The dissertation presents seven separate empirical studies, differing in methodology and research questions addressed. On the one hand it is an important objective of the dissertation to contribute to an increased understanding of the phenomenon of highly cited papers. On the other hand, the thesis attempts to make a contribution to selected issues concerning the applicability and validity of citation indicators. In respect to science policy issues questions are addressed such as: What role can citations/highly cited papers have in order to identify high quality research? How useful are citations given the skewed distributions? The indicator questions are raised at two levels: for assessing the performance of individual publications and for using citation data to characterise a population.

A basic methodological point is that in order assess how robust and reliable is the citation indicator, the nature of its building blocks and their aggregation have to be examined. Highly cited papers represent particularly influential building blocks. Following Rip (1997), a distinction is used between intrinsic and extrinsic validations: Analysing the building blocks of citation indicators will provide more knowledge on what the indicator stands for and may therefore be considered as a kind of intrinsic validation. In extrinsic validations the viewpoint of the scientific community is examined in order to see how the citation counts correspond with the researchers own opinions of scientific contribution. Both types of “validations” are visible in the studies.

Four of the studies deal with intrinsic validation: the study of field delineation, the study of self-citation, the study of the effects of highly cited papers on macro indicators and the study of the structural characteristics of highly cited papers. Three of the studies can be characterised as extrinsic validation: the studies of the relation between bibliometric indicators and peer review and the two studies of authors’ perceptions of citation rates and citation processes.

2.3 Data and methodology

As bibliometric data source I have used the databases NIFU STEP (Norwegian Institute for Studies in Research and Higher Education) has purchased from Thomson ISI. The basic database is the *National Citation Report* (NCR) for Norway, containing bibliographic information for all Norwegian articles (articles with at least one Norwegian author address). Data for each paper include all author names, all addresses, article title, journal title, document type (article, review, editorial, etc.), field category, year by year and total citation counts and expected citation rates (based on the journal title, publication year and document type). In addition, there is a file of all citing papers (i.e. the papers that have cited the “Norwegian” papers) containing the same type of bibliographic data as the source papers. The latter file was used to study self-citations.

In addition, two other bibliometric databases have been used: the *National Science Indicators* (NSI) database containing aggregated bibliometric data at country and field/subfield level and the *Journal Performance Indicator* (JPI) database, containing aggregated bibliometric data at journal level. These databases were mainly applied for the purpose of creating reference standards. Another database located at NIFU STEP that has been applied in some of the studies contains information on the R&D personnel in Norway (age, positions and institutional affiliations, etc.).

In addition to these databases, data for the dissertation were collected through questionnaire surveys. Thus, the dissertation mainly involves two different methods: quantitative bibliometric studies and questionnaire surveys. The quantitative studies have been carried out by using software such as Microsoft Access (the bibliometric data reside in MS Access format). Various search alternatives, queries and programming were applied in these studies. Further description of the methodology of the different studies can be found in the respective articles.

It is a common characteristic of all of the studies included in the dissertation that they are based on Norwegian data. The main reason for this is the limitation of the NCR database located at NIFU STEP. Nevertheless, the studies address general topics. It is therefore appropriate briefly to characterise Norway as a research nation and discuss the implications of using Norway as a case on the conclusions of the thesis.

Norway is a relatively small scientific nation. Less than one percent of the world's total R&D expenditures can be attributed to Norway. In terms of scientific production, approximately 5000 scientific articles are indexed annually by ISI. Norway nevertheless ranks among the 25 largest scientific nations in terms of publication counts. The country has four universities and a few colleges with university status. The institute sector (governmental and private research institutes) is fairly large compared to many other countries. There are also a few "state" colleges where some R&D work is carried out. The industry's share of the total R&D is lower in Norway than in many other Western countries. Furthermore, the country has research activities in a broad range of scientific specialities, but is specialising in research related to its natural resources. The overall average citation rate for Norway is equal to the world average.

Despite its small size, Norway should not be considered as a country on the scientific periphery. The country is highly integrated into the international scientific arena: most of the scientific output is published abroad, and more than a third of the publications show international collaboration. Accordingly, Norwegian researchers collaborate extensively with researchers from other countries. Furthermore, analysing the citations patterns we find that a large majority of the authors citing Norwegian papers are foreigners, suggesting that the results of Norwegian science are mainly used abroad.

In the studies of the dissertation Norway is used as a case for addressing general issues. Because of the high share of international co-authorship among the papers and the relatively low percentage of citations from Norwegian authors, I think the peculiarities of Norway as a research nation should not be given too much emphasis when interpreting the results. On the contrary, representing a small but scientifically integrated country I believe that Norway may be well suited as a case for analysing the various questions concerning citations. Of course, it is still an open question to what extent the findings have general validity. But this is a general methodological question where it is more important to consider that the number of citations and papers analysed in the different studies varies from a few hundred to more than half a million.

3. ARTICLE I. The effect of highly cited papers on national citation indicators.

Aksnes, D. W. & G. Sivertsen

Published in Scientometrics 59(2) (2004): 213-224.

Jointly published by Akadémiai Kiadó, Budapest *Scientometrics,*
and Kluwer Academic Publishers, Dordrecht *Vol. 59, No. 2 (2004) 213–224*

The effect of highly cited papers on national citation indicators

DAG W. AKSNES, GUNNAR SIVERTSEN

Norwegian Institute for Studies in Research and Higher Education (NIFU), Oslo (Norway)

Citation distributions are extremely skewed. This paper addresses the following question: To what extent are national citation indicators influenced by a small minority of highly cited articles? This question has not been studied before at the level of national indicators. Using the scientific production of Norway as a case, we find that the average citation rates in major subfields are highly determined by one or only a few highly cited papers. Furthermore, there are large annual variations in the influence of highly cited papers on the average citation rate of the subfields. We conclude that an analysis of the underlying data for national indicators may be useful in creating awareness towards the occurrence of particular articles with great influence on what is normally considered an indicator of "national performance", and that the common interpretation of the indicator on research policy level needs to be informed by this fact.

Introduction

Citation distributions are extremely skewed. A large part of the scientific articles are never or seldom cited in the subsequent scientific literature. On the other hand, some papers receive a very large number of citations. As an example, among the 75,000 Norwegian ISI-indexed articles published in the period 1981–1998, 40% have never been cited or have only been cited once or twice (counting citations 1981–2002), while ten percent of the papers have received half of the citations. Similar distribution patterns can be found for all countries.

The skewed pattern of citation distributions was identified at an early stage by Price.[1] Although it is now a well-known bibliometric phenomenon,[2] it has not in our view been given sufficient attention, particularly not when considering the large expansion of citation studies the recent years. While there have been intensive methodological debates concerning what is measured by citations, the phenomenon of skewness has mainly been analysed in respect to productivity distributions. When adopted on citation distributions, the perspective has usually been limited to small numbers of articles.

Received October 27, 2003

Address for correspondence:
DAG W. AKSNES
Norwegian Institute for Studies in Research and Higher Education (NIFU)
Hegdehaugsvn. 31, NO-0352 Oslo, Norway
E-mail: Dag.W.Aksnes@nifu.no

0138–9130/2004/US $ 20.00

The focus of this paper is on national citation indicators, which represents a high level of aggregation where one would not expect that only a few articles could influence a general performance indicator. A national citation indicator is basically calculated as the average citation rate of all articles in a given field: The total number of citations in the field is divided by the total number of articles in the same field. On this basis comparisons are made between e.g. other countries or the worldwide average. Thus, the indicator is normally interpreted as an indicator of a country's general performance in the field. This interpretation may, however, be misleading, since it directs the attention away from the fact that a relatively small proportion of highly cited articles may contribute heavily to the overall citedness of a country, and that the majority of articles are seldom or never cited at all. Often, relatively small fluctuations in the citation indicator are considered to be of great importance in policy discussions, but the underlying distribution of citation frequencies is seldom analysed. Furthermore, large annual fluctuations in the citation rate are quite often concealed by using overlapping periods for the calculation, or regression lines for the presentation of trends. For example, in an article by Glänzel[3] this argument is put forward: "Mean citation rates of small and medium-sized countries proved to be liable to annual variations. In order to compensate "random" distortions caused by a "snapshot effect" publications years and citation windows have, therefore, been shifted, and the corresponding counts have been summed up."

On this background we have addressed this question: What effects do highly cited papers have on the national citation indicators? The basic idea is that an examination of how the average citation rate is influenced by highly cited papers may give further knowledge on what the indicator actually indicates. This issue has only been given attention in a few former studies. An analysis of journal impact factors by Moed et al.[4] suggested that for some journals with a high impact factor, the impact was largely determined by a limited number of extremely highly cited articles. Similar conclusions were drawn by Seglen.[2] We will demonstrate that this is even so on the level of national citation indicators. Thus, an analysis of the underlying distribution may reveal other properties than one is normally aware of in the research policy discourse.

The study has been designed as a case study of the scientific production of Norway. The overall national citation indicator for Norway is equal to the world average. In terms of scientific production, the country represents a relatively small nation with a production of less than 5,000 scientific articles that are indexed annually by the Thomson ISI. Still, Norway ranks among the 25 largest scientific nations in terms of publication counts. The country is also highly integrated into the international scientific arena: Most of the scientific output is published abroad, and more than a third of the publications show international collaboration. Norway may, therefore, be well suited as a case for analysing effects of highly cited papers, typifying issues relevant for small, but scientifically integrated countries, or for studies of scientific subfields in all countries.

Methods

The basis for our study is data provided by Thomson ISI. We have applied the two databases: National Science Indicators (NSI), and the National Citation Report (NCR) for Norway. The NSI-database contains aggregated publication and citation counts for all countries in 24 different fields (standard version) and is commonly used for assessing the scientific performance at national levels. Our NCR-database contains bibliometric data on individual articles from Norway. By using the NCR we identified the Norwegian subset of articles within NSI (counting articles, notes, reviews, and proceedings papers). In this way, it has been possible to assess the underlying citation distribution of the macro indicators of NSI. In our study we have excluded fields within the social sciences. The remaining 19 fields within medicine, the natural sciences and technology have been analysed. These fields show large variations in terms of size (number of articles), citation impact and in the range of the citation distributions.

The data have been analysed by using a 5-year citation window. That is, for an article published in e.g. 1981 we have counted the citations to this article in the 5-year period 1981–1985. As basis for the analysis we used the papers published in the period 1981–1994. A 5-year interval has been selected because it is often used in bibliometric analyses and is intermediate in respect to a short-term and a long-term assessment. Since the variability of citedness would be expected to increase with the size of the citation window, it is also sufficient long-term for a distinct polarisation pattern to occur.

A mean citation rate for a particular field has then been calculated as the total number of citations received divided by the number of items published. In this way the mean citation rates for each (publication) year have been estimated. This is a usual procedure in the calculation of citation indicators, although the length of the citation window varies. However, overlapping periods or regression lines are often used instead of single year calculations, particularly when illustrating development over time, as we have seen. In the NSI-database, for example, one of the standard indicators is a 5-year citation indicator using overlapping periods. (That is, an indicator is calculated for papers published in e.g. 1991–1995, 1992–1996 by counting the citations to those papers within the same 5-year windows.)

As a first approach we have calculated the citation indicator for each NSI-field using a 5-year running citation window. Furthermore, we have made different kinds of analyses of the underlying citation distributions, focusing on the high end of the citation range. For each NSI-field we have calculated the percentage of the field citations obtained by the most cited and the five most cited articles for every year.

This method has been applied for all NSI-fields regardless of their size. In our data the smallest field is Computer Science with an average national (Norwegian) production of 18 papers annually, the largest is Clinical Medicine, with an average production of 900 articles annually. These extreme variations in field size will of course be reflected in our results. It should also be added that our method does not adopt any particular criterion for being highly cited. In some small subject categories, papers identified as e.g. "top five" may not be very highly cited. Although these factors may distort interfield comparisons, it has been our intention to show how the effect of a fixed number of papers may vary when using the NSI-categories in bibliometric analyses.

As a case for a particular in-depth study, Neuroscience has been selected. In order to see how the annual citation rates may be influenced by other factors, we have also made a study of the percentage of uncited papers and how this percentage correlated with the annual citation average.

Results

Figure 1 illustrates the skewness in the underlying citation distribution of the NSI-indicator. This example is based on all articles from Norway published 1994–1998 and cited in the same period. It should be clear from this example on the highest aggregated level that a small share of the articles contributed with a high share of the citations.

In our further analyses we used a running 5-year citation window instead of overlapping periods, as described above. In Table 1, we have calculated the percentage of the field citations obtained by the most cited and the five most cited articles for every year. This is a simple and readily understandable expression of the weight that these highly cited articles have in the calculation of national citation indicators.

As we can see, a significant share of the citation impact is due to the effect of a few highly cited papers. In many of the NSI fields, only five single articles among a total output of more than a hundred articles may account for approximately half of the citations. However, we find large differences both at the horizontal and the vertical levels.

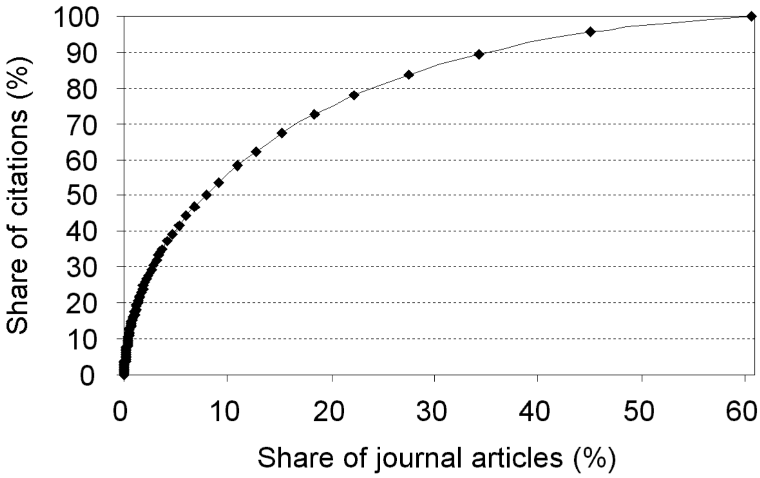


Figure 1. Cumulative contribution of article citations to the NSI-citation indicator (all fields, Norway), 1994–1998 (0-5 year citation window)

On the horizontal level, the variations that correspond with the field sizes (number of articles included) are not unexpected: The few highly cited papers account for the highest shares of the citations in the smallest fields. On the other hand, we find the lowest percentages in the largest field Clinical Medicine. Still, one single paper representing only 0.1% of the production in Clinical Medicine may account for up to 6% of the citations. In the much smaller field Geosciences (average production of 125 papers annually) the percentage is almost as low as in Clinical Medicine and varies between 4 and 6 %. This shows how the degree of polarisation may vary between fields. In some fields we find a strongly skewed distribution, while in other fields it is less strongly skewed. One reason is that the size of the literature affects how many citations a highly cited paper can get. When the population of papers is large (there is a large number of potential citers) and thus much greater probability for extremely highly cited papers to occur.[5] Similarly, the average citation impact of the field may correspondingly influence on this probability. Thus, variations on the horizontal level may also be explained by differences in underlying citation distributions.

Table 1. The percentage of the field citations obtained by the most cited article, by NSI fields/years (5-year citation window)

Publ. year	AGD	ASD	BID	CHD	CLD	CSD	EGD	EVD	GED	IMD	MBD	MCD	MSD	MTD	NED	OTD	PHD	PLD	PMD
1981	10	38	4	8	5	43	14	8	6	16	8	6	14	18	11	28	5	3	5
1982	8	29	3	3	2	28	17	9	4	8	19	6	24	10	5	22	4	6	10
1983	8	19	5	3	2	31	15	7	6	10	14	6	25	24	21	8	5	13	7
1984	12	10	3	2	3	17	17	9	5	9	8	22	11	13	25	13	6	3	7
1985	6	31	5	4	1	21	23	16	5	12	15	16	19	17	6	13	7	2	8
1986	8	23	5	2	2	32	10	6	5	12	15	11	24	19	9	11	8	3	7
1987	18	21	4	3	6	55	13	15	5	7	13	13	15	11	21	19	8	3	6
1988	10	12	5	6	2	25	9	15	6	8	22	8	12	10	12	12	7	3	7
1989	11	22	7	6	5	34	10	5	4	7	11	17	8	13	4	31	23	2	7
1990	18	27	6	4	2	21	8	5	5	7	12	8	9	10	4	27	6	2	8
1991	9	16	5	3	3	38	7	7	3	7	12	7	14	17	8	11	5	3	21
1992	7	13	6	4	2	32	4	15	4	9	10	6	11	19	14	10	2	5	9
1993	11	10	3	3	2	20	6	8	3	4	7	7	16	7	5	16	7	3	20
1994	10	20	6	3	3	22	10	5	5	7	8	18	8	12	7	8	4	3	10

The percentage of the field citations obtained by the 5 most cited articles, by NSI fields/years (5-year citation window)

Publ. year	AGD	ASD	BID	CHD	CLD	CSD	EGD	EVD	GED	IMD	MBD	MCD	MSD	MTD	NED	OTD	PHD	PLD	PMD
1981	29	79	15	17	11	86	45	31	28	36	29	21	47	50	30	65	23	12	19
1982	30	78	13	12	10	72	56	29	17	24	43	22	60	44	21	58	19	16	21
1983	30	70	14	13	7	86	41	25	27	29	40	23	69	54	42	36	20	19	23
1984	44	43	11	9	8	71	35	33	20	26	33	46	46	43	47	39	22	13	24
1985	30	78	16	15	5	84	49	36	21	34	34	32	54	39	24	37	23	9	30
1986	29	66	17	10	7	75	34	22	20	38	44	34	57	47	23	45	24	11	26
1987	47	59	13	13	11	90	34	39	19	26	40	41	47	43	47	56	25	11	27
1988	32	50	13	17	7	65	33	31	20	25	50	32	43	40	26	48	23	11	28
1989	34	59	19	18	12	85	37	18	16	30	35	43	32	35	19	83	33	9	27
1990	66	71	16	14	10	67	34	21	16	25	36	33	38	43	18	62	20	8	28
1991	37	49	16	13	8	76	24	25	12	31	37	27	39	56	23	47	20	12	34
1992	23	50	16	12	6	77	20	34	13	22	28	25	37	50	29	29	11	13	29
1993	37	42	14	10	9	69	21	25	14	18	24	26	45	28	16	45	19	12	43
1994	32	42	13	12	9	65	24	18	15	23	29	34	33	46	23	29	13	8	33

AGD: Agricultural Sciences
ASD: Astrophysics
BID: Biology & Biochemistry
CHD: Chemistry
CLD: Clinical Medicine
CSD: Computer Science
EGD: Engineering
EVD: Ecology/Environment
GED: Geosciences
IMD: Immunology
MBD: Molec. Biology & Genetics
MCD: Microbiology
MSD: Materials Science
MTD: Mathematics
NED: Neuroscience
OTD: Multidisciplinary
PHD: Physics
PLD: Plant & Animal Science
PMD: Pharmacology

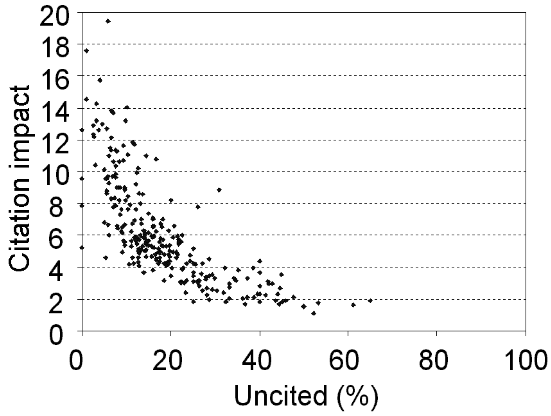


Figure 2. Distribution of citation impact and uncitedness
(plotting of annual scores for 19 NSI-categories 1981–1994, using a 5-year citation window)

The focus of our further analysis was, however, directed at the vertical variations: Changes over time within subfields. These annual variations were striking and needed further explorations. As an example, the most cited article in Engineering accounts for between 4% and 23% of the field citation impact, depending on the year of measurement. Thus, the presence or absence of one particularly highly cited paper may drastically change the average citation impact of the field, and thereby the national citation indicator. In one instance in Physics, we found that the occurrence of one highly cited paper influenced the national citation indicator to a high degree without having any relevance as an indication of 'national performance'. A highly cited CERN-article accounts for 23% of all citations to Norwegian articles in Physics in a particular year, while this share fluctuates between 2% and 8% in all other years. This article has an extremely large number of authors from many countries. One institutional address in Norway significantly affects the national citation indicator for Physics that year.

We proceeded to study how the annual variations in the average citation impact correlated with changes in the annual importance of highly cited papers. The correlation was somewhat weaker than expected. Thus, the annual variations in the national citation indicator are being determined by other factors as well. For example, there may be large annual variations also at the other ends of the citation distributions, such as in the percentage of uncited/poorly cited papers or in the number of moderately highly cited papers. In Figure 2 we have analysed the opposite 'tail' of the skewed distribution, that is, the number of non-cited articles per year. For each NSI-subfield/year we have calculated the citation impact and the percentage of uncited papers (annual scores for 19 NSI-categories 1981–1994, using a 5-year citation window), and the results have been plotted.

As we can see, the two measures show a strong, negative correlation. This is not surprising, since the two measures are dependent upon each other (see also Ref. 4). Thus, when the citation impact is high, the percentage of uncited articles is very low. Similarly, when the citation impact is low, the percentage of uncited papers is generally high. We do, however, again find significant horizontal and vertical differences. These differences can be explained by other factors, e.g. by annual variations in the effects of highly cited papers, as shown above. In conclusion, our analysis of correlations showed that one should look at both extreme ends of the citation distribution in order to find the factors that to a higher degree than "the average article" may influence the citation indicators based on average measures.

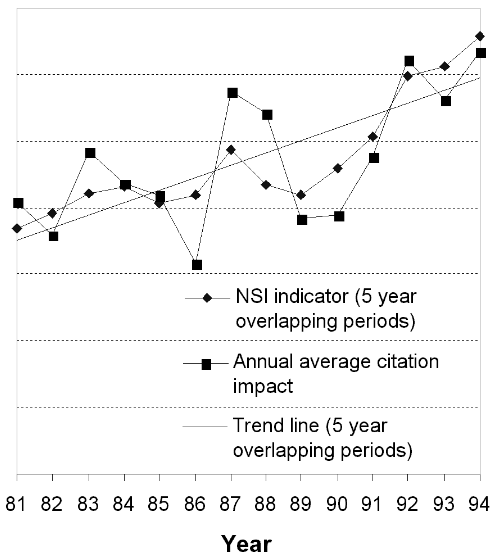


Figure 3. Citation indicators for Neuroscience (NED): NSI indicator (5-year overlapping periods), annual average citation impact (5-year citation window), and trend line*

*) The indicators have been adjusted to allow for comparisons: The annual average citation impact has been adjusted to the scale of the NSI indicator by multiplying it with the sum of the annual NSI indicator impacts divided by the sum of the annual average citation impacts. The NSI indicator has been plotted the first year of the periods (e.g. period 81-85 is plotted in 81

Table 2. The presence of highly cited articles in a medium sized NSI-field, Neuroscience (NED) (5-year citation window)

Publication year	Total number of citations in the field	Total number of articles in the field	Avarage citation impact	Number of citations, most cited article	Percentage of citations obtained by the most cited article	Number of citations, 5 most cited articles	Percentage of citations obtained by the 5 most cited articles
1981	569	68	8.4	65	11 %	173	30 %
1982	786	107	7.3	39	5 %	164	21 %
1983	872	88	9.9	186	21 %	369	42 %
1984	1,209	136	8.9	305	25 %	563	47 %
1985	883	103	8.6	50	6 %	208	24 %
1986	798	124	6.4	69	9 %	186	23 %
1987	1,372	117	11.7	290	21 %	647	47 %
1988	1,181	107	11.0	145	12 %	302	26 %
1989	995	127	7.8	44	4 %	192	19 %
1990	1,176	148	7.9	52	4 %	215	18 %
1991	981	101	9.7	83	8 %	226	23 %
1992	1,800	142	12.7	250	14 %	530	29 %
1993	1,468	128	11.5	67	5 %	234	16 %
1994	1,997	154	13.0	149	7 %	466	23 %

Annual fluctuations in the average citation impact of a particular field are often "hidden" in bibliometric reports by the use of overlapping citation windows or trend lines. This effect is showed in Figure 3, in which we have compared different citation indicators for Neuroscience, including the standard NSI indicator involving overlapping 5-year periods. As a source of reference, we have also included in Table 2 to show the actual figures of the effects of highly cited papers in this field. Neuroscience represents a medium sized field with more than 100 papers annually. Still, as we can see, the most highly cited paper accounted for up to 25% of the citations.

One important issue needs, however, to be addressed. Our study has been based on the standard way of allocating articles: A paper is attributed to a country if the paper carries at least one author address of that country. This is also the method applied in the National Science Indicators (NSI). An alternative method would be to use fractional counting, which credits the authors of multiple authored papers their fractional contribution (for example applied in the US science indicator reports[6]). Because highly cited papers tend to have many authors, often involving international collaboration,[7] this method might reduce the influence of the few highly cited papers. In particular, this would have been the case in the example with the CERN-article in Physics, discussed above. In order to analyse this question further, we made a test for the year 1994. We calculated the percentage of the field citations obtained by the five most highly cited articles (cf. Table 1) using fractional counting. In several cases this reduced the relative influence of the highly cited papers, but the reduction was generally rather limited. Somewhat surprisingly, the opposite effect could also be identified for several of the subfields. We therefore conclude that fractional counting would not significantly change the pattern identified in our study, although this method generally might be considered as better or more "fair" than the standard method.

Discussion

Two main findings result from our study: (1) The average citation rates in national subfields are to a large extent determined by only a few highly cited papers. (2) There are large annual variations in the influence of the few highly cited papers on the average citation rate of the subfield.

It is well known that citation distributions are extremely skewed and that only a minor part of the articles are cited near the sample mean. Our study shows that the phenomenon has implications even for the interpretation of indicators on a country level. Analyses of national performances presented for example in science and technology indicator reports are still usually based on average citation scores only (or total number of citations).[6,8] Thus, it is this indicator that is presented for a wider science policy public, and functions as the basic premise for particular conclusions regarding the scientific performance of a country. On the basis of our findings one might raise questions concerning the common interpretation of the indicator as a measure of the performance of a nation in general. Is the average citation impact of all articles from a nation an indicator of its 'normal' scientific activity, or is it, rather, an indicator of the capability of a nation in producing highly cited papers?

Our study suggests that observed variations in the citation indicator should lead to a closer attention to particular problems or events in specific research environments (the reasons for the variations). Because of the great variability in citedness, the average might disguise interesting underlying patterns. For example, using a high average score to conclude that a nation is performing well in a particular subfield could be misleading. It might possibly be the case that this high score is a result of one group producing particularly highly cited papers, while the production of all other scientists within the field is being moderately or poorly cited. In that case, the conclusion would rather be that the country has one leading group within the field, but is generally not performing well.

One might consider the median to be a better measure of what is the "typical" value. In fact, in other contexts dealing with data containing outliers (e.g. personal incomes) the median is often used instead of the mean as a measure of central tendency. In citation studies the median is still not very suitable as an indicator because it will bring all countries near to the same score (e.g. 1 citation or 2 citations), and also because it does not reflect characteristics and differences at the high end of the citation distributions. In addition, the calculation of the median presupposes knowledge of the underlying distribution, which is not available from standard bibliometric databases such as the NSI.

It appears that an extreme skewness is a statistical regularity characterising all citation distributions. This pattern can be found at a variety of levels, e.g. for articles of a journal, for countries, scientific fields and even for individual authors.[2,9] For example, Seglen found the distribution of citedness within scientific fields to approach linearity in a double-log plot, while the variability between articles in a journal showed a less skewed, semilog linearity. On this basis it can be argued that it is still legitimate to use the citation average as basis for comparisons as long as the samples analysed show similar skewed distributions. However, this study has shown that other indicators than the mean should be used in order to get a richer picture. Such indicators might include the number of highly cited papers or the percentage of uncited papers. Highly cited papers have, in fact, been found useful as indicators in some research evaluations,[10–13] and recently highly cited papers were included as indicators in the benchmarking project of the European Commission.[14]

As we have seen, overlapping periods and regression lines are often used in order to compensate for what is considered as "random" variations in the annual citation impact. Nevertheless, is not obvious why such variations should be considered as "random". This paper has shown that depending on the purpose, it may be relevant to consider that the citation impact of a nation in a particular field may not be very stable, but shows large annual variations, and that this is normally the case. The reasons for these variations should rather be analysed in further detail, than disguised by abstract measures. Indicators that identify highly cited articles may be important in this context.

Conclusions

Whenever aggregated citation indicators are used to assess the scientific performance of a nation, the underlying citation distributions are seldom analysed. These distributions are always extremely skewed. An analysis of the underlying data for national indicators may, therefore, be useful in creating awareness towards the occurrence of particular articles with great influence on what is normally considered an indicator of "national performance" in general. Normal science is seldom or never cited, and in this respect all countries resemble each other. Accordingly, the measurement of citations on the level of national indicators should be able to capture deviations from the "normal science".

*

An earlier version of this paper was presented at the 8th ISSI conference, Sydney, July 2001.

References

1. D. J. DE SOLLA PRICE, Networks of scientific papers, *Science*, 149 (1965) 510–515.
2. P. O. SEGLEN, The skewness of science, *Journal of the American Society for Information Science*, 43 (1992) 628–638.
3. W. GLÄNZEL, Science in Scandinavia: A bibliometric approach, *Scientometrics*, 48 (2000) 121–150.
4. H. F. MOED, T. N. VAN LEEUWEN, J. REEDIJK, Towards appropriate indicators of journal impact, *Scientometrics*, 46 (1999) 575–589.
5. E. GARFIELD, The use of journal impact factors and citation analysis for evaluation of science, Paper presented at the *41st Annual Meeting of the Council of Biology Editors*, Salt Lake City (1998), available at: http://www.garfield.library.upenn.edu/papers/eval_of_science_CBE(Utah).html
6. NATIONAL SCIENCE BOARD, *Science and Engineering Indicators – 2002*. National Science Foundation, Arlington 2002.
7. D. W. AKSNES, Characteristics of highly cited papers, *Research Evaluation,* 12(3) (2003).
8. EUROPEAN COMMISSION, *Third European Report on Science & Technology Indicators 2003,* Brussels 2003.
9. P. O. SEGLEN, Causal relationship between article citedness and journal impact, *Journal of the American Society for Information Science*, 45 (1994) 1–11.
10. B. R. MARTIN, J. IRVINE, Assessing basic research: Some partial indicators of scientific progress in radio astronomy, *Research Policy*, 12 (1983) 61–90.
11. R. PLOMP, The highly cited papers of professors as an indicator of a research group's scientific performance, *Scientometrics*, 29 (1994) 377–393.
12. B. R. MARTIN, The use of multiple indicators in the assessment of basic research, *Scientometrics*, 36 (1996) 343–362.
13. R. J. W. TIJSSEN, M. S. VISSER, T. N. VAN LEEUWEN, Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54 (2002) 381–397.
14. EUROPEAN COMMISSION, *Key Figures 2001. Special Edition. Indicators for Benchmarking of National Research Policies*. Brussels 2001.

4. ARTICLE II. Characteristics of highly cited papers.

Aksnes, D. W.

Published in Research Evaluation 12(3) (2003): 159-170.

Research Evaluation, volume 12, number 3, December 2003, pages 159–170, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Quality assessment

Characteristics of highly cited papers

Dag W Aksnes

Highly cited articles are very different from 'ordinary' cited articles. Typically, they are authored by a large number of scientists, often involving international collaboration. The majority of the papers represent regular journal articles (81%), although review articles (12%) are over-represented compared to the national average. The citation curves of highly cited papers follow a typical pattern of rise and decline. However, different types of citation curves can be identified, reflecting possible differences in the cognitive function of the articles. Highly cited papers typically obtain citations from a large number of different journals and from papers representing both close and remote fields. However, this pattern is not very different from the average distribution for all papers. We discuss how the findings can be explained by introducing a conceptual distinction between quality dynamics and visibility dynamics.

Dag W Aksnes is at the Norwegian Institute for Studies in Research and Higher Education (NIFU), Hegdehaugsveien 31, NO-0352 Oslo, Norway; e-mail: Dag.W.Aksnes@nifu.no

The author would like to thank Arie Rip for valuable comments and suggestions on an earlier draft of this paper.

CITATION DISTRIBUTIONS ARE extremely skewed. The large majority of the scientific papers are never or seldom cited in the subsequent scientific literature. On the other hand some papers receive an extremely large number of citations (Aksnes and Sivertsen, 2004); Seglen, 1992). During the last decade there has been an emerging interest in using highly cited papers as indicators in research assessments. One reason for this is the increasing focus on scientific excellence in science policy (Van Raan, 2000). In this context, highly cited papers have been regarded as potential candidates for identifying and monitoring 'excellent' scientific research. Lately this was shown in a benchmarking study from the European Commission in which highly cited papers were used as indicators for comparing the research performance of the EU countries (European Commission, 2001). Highly cited papers have also been applied as indicators in case studies of research groups (e.g. Martin and Irvine, 1983), and an explorative study by Tijssen *et al* (2002) concluded that highly cited research papers do represent useful indicators for identifying 'world-class' research.

However, the application of citation indicators is controversial and involves important elements of uncertainties. It is not clear what one is 'measuring' using highly cited papers and how these measures relate to various aspects or conceptualisations of scientific excellence. Against this background it is urgent to look further into the phenomenon of highly cited papers. Especially in small and peripheral countries, where the need to be selective is largest, the citation indicator is more uncertain than in core countries. We will therefore analyse one such country: Norway.

The focus of the study lies in the bibliometric and quantitative characteristics of highly cited papers. There have been a few earlier studies of this issue.

Generally, highly cited papers have been found to be very different from 'ordinary' cited papers. For example, it has been shown that highly cited articles tend to be authored by many researchers and are typically the result of international collaboration (Glänzel *et al*, 1995). The purpose of this study is to examine various findings concerning the bibliometric characteristics of highly cited papers, assessing their generality in respect to a sample of highly cited papers in Norwegian science. Former studies are limited in scope and numbers and we will also address new hypotheses on this topic. From a research policy perspective it is of particular interest to assess these hypotheses because they are involved in the interpretation of citation data for research evaluations.

In addition to analysing how the highly cited articles in Norwegian science differ from an 'average' publication, we will also differentiate among the highly cited papers. That is, we will analyse to what extent there are internal differences among the papers. Such a study can contribute to a further understanding of the cognitive as well as the social factors involved in frequent citations. By introducing a distinction between quality dynamics and visibility dynamics in the latter part we will discuss how our findings can be explained. Through a case study of Norwegian science in this way we will make further contributions to knowledge of highly cited papers.

Methods

There are various definitions of what counts as a highly cited article. Basically two different approaches can be identified, involving absolute or relative thresholds. Sometimes a fixed threshold is used as a definition; for example articles cited more than 400 times. The consequence of using a fixed number of citations as a limit is, however, a predominance of articles from highly cited fields. There are large disciplinary differences in the average citation frequencies, and similar differences can be found at the high end of the citation range (Aksnes and Sivertsen, 2004). For example, a highly cited article in mathematics may be considered only as an ordinary cited publication in molecular biology. A relative standard is, therefore, often adopted instead. Such a selection method identifies the most highly cited papers within each field.

In this study we have used a relative standard. That is, a paper has been considered as highly cited if it has received more than a certain multiple of the citations of the average paper within the scientific subfield. By this we have adjusted for the large differences in citation rates between different subfields. The concept of highly cited paper is thus based on variable and field specific standards.

The basis for our study is data provided by the Institute for Scientific Information (Thomson ISI). We have applied the two databases — National Science Indicators (NSI) and the National Citation Report (NCR) — for Norway. The NSI database contains aggregated publication and citation counts for all countries in 105 different fields (de luxe version) and is commonly used for assessing scientific performance at national levels. Our NCR database contains bibliometric data on individual articles for Norway (that is, publication with at least one Norwegian address). By using the NCR we identified the Norwegian subset of articles within NSI (counting articles, notes, reviews and proceedings papers). We applied the 2001 editions of these ISI products, with data covering 1981–2000. We excluded articles from the humanities and the social sciences because of the limitations of the ISI indicators in these areas.

Selection criteria

A small scientific nation such as Norway has a rather limited annual production of papers being very highly cited. For that reason alone, we have analysed the scientific production over a longer period of time: the 15-year period 1981–1996.

We applied a five-year citation window as basis for our analyses. That is, for an article published in e.g. 1991 we counted the citations to this article in the five-year period 1991–1995. A five-year interval was selected because it is often used in bibliometric analyses and is intermediate with respect to a short- and a long-term citation window. Since the variability of citedness is expected to increase with the size of the citation window, a five-year interval is sufficient long term for a distinct polarisation pattern to occur. By applying such a methodology, articles with a very slow or 'delayed' citation growth, may, however, be missed — but such articles are rather unusual.

A publication is been considered as highly cited if the number of citations received is more than a certain multiple of the mean citation rate of the particular subfield. Such a method of selection has similarities to the method applied by Glänzel *et al* (1995) in a study of highly cited papers in physics. However, here an additional criterion was introduced by requiring the papers to be cited above a certain minimum level.

Due to practical limitations in the availability of data, the field averages applied for comparison have been calculated as five-year overlapping periods (and not as standard five-year citation windows). This means that for articles published in e.g. 1981 we have compared the citation rates with the NSI-indicator for the period 1981–1985 (calculated as the inverse ratio between the number of papers published in the five-year period and the citation counts to those papers within the same five-year window). Similar calculations have been carried out for all the years 1982–1986 … 1996–2000. Although the NSI indicator used for comparisons differs somewhat in method of calculation, this is unlikely to have any practical implications for identifying the sample.[1] The important point is that both the field and time differences in citation rates are taken into account by this way of calculation.

The basic principle in determining the threshold value has been that the selected articles should be cited much more than the average paper of the subfield (‘compared on equal terms’). In addition the sample identified should be manageable from a practical point of view, meaning that the number of papers should not be too large (or too small) for carrying out the different surveys. Using the references standard described above (running citation window) we selected a ‘score value’ of 17. That is, a publication has been considered as highly cited if the number of citations received during the time period is at least 17 times the mean NSI-citation rate in the particular subfield/year (this would correspond to approx. eight or nine times the ‘real’ field average in 1996). The particular threshold of selection is, of course, somewhat arbitrary. The sample of highly cited papers will change continuously if the criterion is being changed. Still, the publications represent the very top in their fields in the scientific production of Norway during a 15-year period.

Identifying the sample

As a first step we identified the Norwegian subset of articles within NSI, 1981–1996, a total of 49,945 publications. Then we added the field codes to the articles. Because some of the articles are classified in more than one subfield, this increased the number of units to 63,565. We then excluded articles within the social sciences and the humanities (social sciences, general, arts and humanities, education, economics and business, and law) as well as articles with missing field codes. We were then left with a sample of 58,616 articles. Excluding ‘double counts’ this corresponded to 46,849 unique articles. These publications represent the overall national publication set from which comparisons have been made.

We then calculated the citation counts to the articles within a five-year window and obtained data on the average citation rates of the subfields. A ‘score value’ for each article was calculated by dividing the number of citations on the field average. Using the score value of 17 as a selection criterion we were left with a set of 346 highly cited articles. Of these, 49 articles appeared twice, reducing the number of unique articles to 297. The papers represented a variety of different scientific subfields, although the share of the production identified as highly cited varied significantly among the subfields.

Results

Below we present the results of the study. The text is organised in two main parts. In the first part we consider particular issues/findings concerning highly cited papers drawing on earlier studies. We assess these explicitly as hypotheses and discuss to what extent they are confirmed/falsified by our results. In the second part we present other patterns and possible explanations. Here the style of analysis is exploratory. Each part is divided into section focusing on particular hypotheses/issues.

Hypotheses

Highly cited papers are typically authored by a large number of scientists The number of authors is often used a measure of the extent of scientific collaboration. One basic hypothesis in respect to highly cited papers is that these typically are the results of collaboration between many researchers. Former studies have shown a general correlation between citation counts and number of authors (e.g. Peters and Van Raan, 1994). Such a correlation is also found in the population of Norwegian articles (Table 1). A publication with four authors on average obtains twice as many citations as a publication with only one author.

In our analysis of the highly cited articles we find that 8.9 scientists author the average paper. However, there are large differences within the sample. Four papers were authored by more than 200 authors (ISI only indexes a maximum of 200 authors in this database) and 23 (8%) of the papers were authored by only one person. Furthermore, the paper representing the median was written by ‘only’ four scientists. The number of authors of the highly cited papers is, nevertheless, significantly higher than in the overall publication set, represented by an average of 3.7 authors. Our results thus verify the

Table 1. The relation between number of authors and average citation rates* for Norwegian scientific papers, 1981–1996

Number of authors	Average citation rate	Number of papers
1	3.8	8,990
2	5.3	11,995
3	6.0	10,073
4	7.2	6,815
5	8.9	3,839
6	10.9	2,129
7	12.7	1,114
8	14.8	572
9	14.7	328
10	17.0	193
11-20	21.6	467
20-50	17.8	177
>50	20.1	157

Note: * Within a five-year period

Our results verify the hypothesis that highly cited papers typically involve more collaborative research than what is the normal or average — but only at an aggregated, general level

hypothesis that highly cited papers do typically involve more collaborative research than what is the normal or average — but only at an aggregated, general level.[2]

These differences should be compared with the general correlation between citation counts and number of authors. The reason for this correlation is a matter of discussion. Generally speaking, papers with many authors will benefit from particular mechanisms that may increase the citation counts, such as effects caused by many potential self-citers and an enhanced dissemination in the research communities through the personal communications of many authors. Still, in another study we showed that self-citations contributed only to a minor part of the overall increase in citation rates that are found for multi-authored papers (Aksnes, 2003). Furthermore, independent studies have found collaboration to increase the quality of the research. For example, using 'non-citational' measures, Lawani (1986) found that the number of high-quality research papers significantly correlated with the number of authors. In other words, there are strong indications that in terms of impact and research quality it pays to collaborate. It is, therefore, not unexpected that the highly cited papers are authored by larger number of authors than what is the average. Some of these papers are likely to be the results of great research efforts both in terms of manpower and economics.

Highly cited papers do typically involve international collaboration Former studies have shown that highly cited papers are characterised by an extensive element of international co-authorship. For example, Glänzel *et al* (1995) found that 41.5% of highly cited European physics papers were published as the result of international collaboration. This rate of internationalisation was much higher than in the total population of publications. A less pronounced tendency was found in a study of highly cited German papers (Glänzel and Czerwon, 1992). Here, 26% of the papers were the results of international collaboration.

We examined to what extent this pattern also could be identified in our sample. The analysis showed that 186 of the highly cited papers (63%) were co-authored by researchers working in other countries. In comparison, 29% of the papers in the overall publication set involved international co-authorship. The co-authorship of 106 of the highly cited papers was bilateral, and of 33 was trilateral, while 47 papers were co-authored by scientists from four or more countries. In the most extreme case, scientists from 22 different nations together wrote a paper. This means that 36% of the highly cited papers were the result of bilateral projects; a share corresponding closely to the share of papers resulting from unilateral projects (37%), while 27% of the papers resulted from multilateral projects.

The hypothesis concerning international collaboration is, thus, clearly verified in our study. We find many more international contributions among highly cited papers than in the total set of publications. Although Norway generally co-operates strongly with foreign scientists, this tendency is much more pronounced when it comes to the highly cited papers. In fact, when 63% of the papers involve international co-authorships, the concept of 'Norwegian highly cited papers' appears rather problematic. At this level delimitations by countries may be difficult to justify — unless correction are being made for international co-authorships.

The results may indicate that in order to produce very high-impact research it is almost a requirement for Norwegian scientists to collaborate with foreign scientists. One possible reason for this is that it is difficult to obtain powerful groups within a small country. However, as we have seen, this pattern is also found in other studies of highly cited papers — and may be exemplified through large-scale projects such as the human genome project and the particle physic research at CERN.

Highly cited papers are mainly present in high-impact journals It is a statistical truism that highly cited papers tend to be published in high-impact journals (journals with high average citation rates). In order to analyse to what extent this was the case in our sample, we collected the impact factors of the present journals, using the Journal Performance Indicator database (JPI). All except three papers were published in journals to which an impact factor could be assigned. We then compared the impact factor with the average for the field to which the journal was assigned using the NSI-database. In both cases we used a running five-year window.[3] Using this method an index number was calculated for all publications. In order to analyse how the distribution of the highly cited papers differed from the average, we made similar calculations for the overall national publication set.

The results are shown in Figure 1. As we can see, the highly cited papers have a very different distribution from what is the 'normal' Norwegian distribution. 20% of the highly cited papers appear in journals with a very high relative journal impact factor (that is, in journals in which the impact factor is more than four times higher than the average citation.

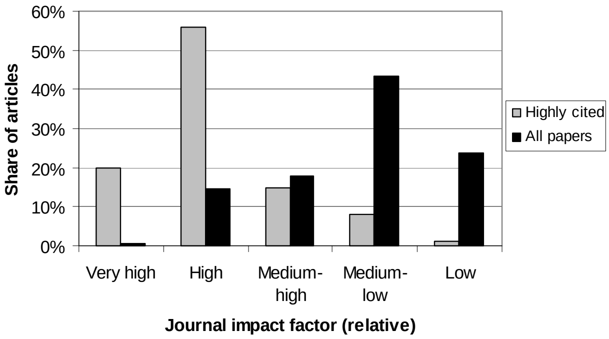


Figure 1. Distribution of articles in different journal categories, highly cited papers vs. all papers, Norway 1981–1996.

Notes: Very high (index >4); high (index 1.5–4); medium-high (index 1–1.5); medium-low (index 0.5–1); low (index < 0.5).
N: highly cited 343; all 54,931. Some journals are assigned to more than one subfield, causing multiple counts.

Similarly, 56% of the highly cited papers appear in journals with a high relative impact factor (1.5–4 times the average). In comparison, only 15% of the total population of Norwegian papers appear in this journal category. Only 9% of the highly cited articles were published in journals with an impact factor below the field average. In contrast, 67% of the Norwegian article production appears in this category of journals cited below average.

Although these results may not seem unexpected, one should notice that 9% of the highly cited papers appear in poorly cited journals. Thus, in order to become highly cited it is not a necessary condition to be printed in high-impact journals or journals of high prestige. This and the skewness in the citation distribution that can be found within most journals indicate that contents are important determinants in citation rates. The journal address represents a contributory factor, but here there is also a choice effect. When scientists think they have something important on their hands (which will qualify for high citation) they will submit the paper to a high-status journal (which usually has high-impact factors). So there is also a content-status effect derived from author's choices.

Review articles are over-represented among highly cited papers Several studies have shown that review articles are over-represented among highly cited papers (Glänzel and Czerwon, 1992). This pattern was also found in our study (Table 2): 12% of the highly cited papers were classified as review articles, while only 2% of all papers represented such articles. In contrast, the share of notes and proceedings papers was lower among the highly cited papers than in the total population of papers. Thus, in terms of review articles, notes and letters, the set of highly cited papers differs in composition from the overall national publication set.

One kind of criticism towards the use of citations as indicators has concerned review articles (e.g.

Table 2. Distribution of articles on different publication types, highly cited papers vs. all papers, Norway 1981–1996

	Regular articles	Reviews	Notes	Proceedings
Highly cited papers	81%	12%	3%	4%
All papers	84%	2%	5%	9%

Seglen, 1997a). Although such articles usually do not contain any new material they are often much cited. Accordingly, they invalidate the use of citations as indicators in research assessments. In our sample the large majority of publications are still not review articles. Thus, the 'review-effect' must be considered as moderate at the high end of the citation range, even if in individual cases it may make a difference of research assessment.

Highly cited papers are mainly cited by foreign scientists By analysing the citing papers we were also able to assess from which countries the citations were received from. One basic hypothesis is that these papers are very international and are being cited mainly by foreign scientists. In order to analyse this issue we identified the countries the citing scientists represented, using the information in the address field of the citing papers. For each citing article the country counts were fractionalised according to the total number of addresses in the paper. A similar analysis was carried out for the overall national publication set.

The results are shown in Figure 2, which shows that 7% of the citations to the highly cited papers come from scientists working in Norway, while the corresponding share in the overall national subset is 19%. For both groups of articles the share of citations from other Nordic researchers is approximately 8%. However, compared to an average paper, the highly cited papers are much more cited by US scientists (26% vs. 33%) and by European scientists (28% vs. 33%).

The results confirm the hypothesis that the highly cited papers are more cited by foreign scientists.

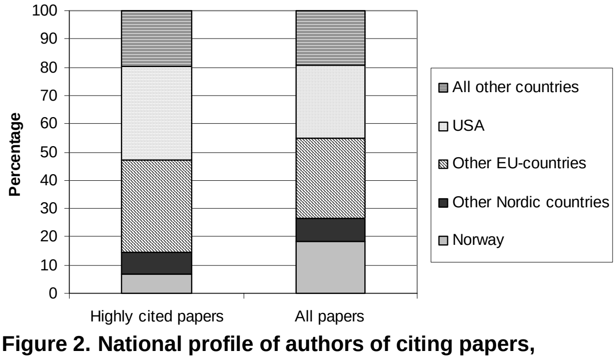


Figure 2. National profile of authors of citing papers, Norway 1981–1996

These papers may, therefore, be considered as particularly important conveyors of scientific knowledge transfer. The analysis also shows that Norwegian science is generally highly integrated into the international scientific arena: only 19% of the citations come from Norwegian scientists (however, Norway contributes only to 0.7% of the articles worldwide). The argument from these findings obviously depends on the size of the nations. For a larger scientific nation like Germany, the share of domestic citations would be much higher.

The share of self-citations is very low for highly cited papers Author self-citations account for a relatively large share of all citations. Particularly when citations are used for assessing scientific impact, these citations are often treated as problematic (MacRoberts and MacRoberts, 1989; Seglen, 1997a). In this study we also identified the number of self-citations for each of the highly cited articles (counting all citations from year of publication through 2000). For each article we calculated the number of citing papers that represented author self-citations, using the criterion that at least one author (first author or co-author) is also an author (first author or co-author) of the citing paper.

In another study we have shown that poorly cited papers generally have the highest self-citation shares (Aksnes, 2003). Therefore, the hypothesis is that the share of self-citations among highly cited papers is relatively low, because high citation counts cannot easily be obtained through self-citations. This is due to the fact that there are practical limits for how often the author(s) can possibly cite the paper.

The results show that on average 15% of the citations represented self-citations. Still, there were large individual differences. In absolute terms, a paper receiving 136 self-citations (among a total of 237 citations) represented the most extreme case. In comparison the average share of self-citations for all Norwegian papers amounted to 21% (29% using a five-year window). Thus, the findings verify our hypothesis, although the share is not much lower than what is the average.

Other patterns and possible explanations

Highly cited papers age slowly

In his pioneering works, De Solla Price analysed various aspects of citation distributions. He realised that although a paper not cited one year may well be cited next year, and an article highly cited in one year may not be heavily cited subsequently, the distributions show strong statistical regularities. In most cases the future citation history of a paper can be determined on the basis of the citations received during the first period (Price, 1965; 1976).

The question of the life-cycle curves of highly cited papers has also been addressed in our study. In

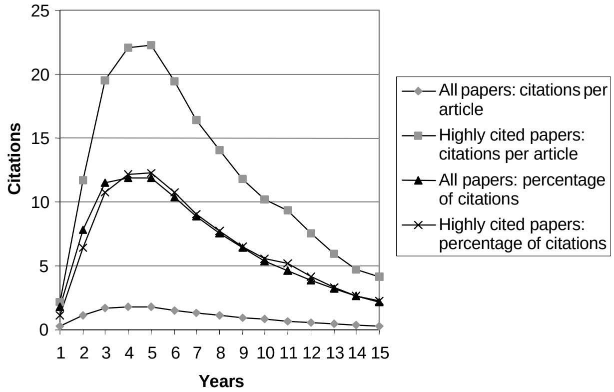


Figure 3. Citations* received vs. time following publication
Note: * The citation counts have been scaled up according to year of publication.

order to analyse this question we made a diachronous citation study. We traced the total number of citations received per year for each document and grouped the citations according to number of years after publication. In order to analyse the life cycle for more than five years, the citation counts were scaled up for an additional five- to ten-year period (depending on year of publication). A corresponding analysis was also carried out for all papers (the national subset).

The results are shown in Figure 3. It is generally known that citations follow a typical pattern of rise and decline. An article is poorly cited the first year, reaches a citation peak a few years after publication and then shows a slowly decreasing pattern of citedness the following years. This pattern is clearly illustrated in the citation curve of the highly cited papers. On average these papers reach a top of 22 citations in the fourth and fifth years following publication (that is three to four and four to five years, respectively, after publication). In comparison, the average for all papers is a maximum of 1.78 citations at identical time periods. Calculated as the percentage of all citations received during the 15-year period, the highly cited papers do not differ much from the average. Both groups of papers receive a maximum of 12% of their citations during each of these peak years. Still, the highly cited papers age slightly more slowly and peak a bit later than the other papers, but the differences are only marginal.

The results of this analysis may seem surprising. It is remarkable that in relative terms the highly cited and the average papers follow an almost identical citation cycle. One factor to be taken into consideration is that citation distributions are highly skewed. The median paper is hardly cited at all, and a few highly cited papers may contribute considerably to the citation mean and thus to the shape of the citation cycle for the average paper (our sample of highly cited papers contributed to 8% of the citations received by the subset of Norwegian publications). However, the highly cited papers have visibility over a long period: 15 years after publication the highly cited papers on average receive 4.2 citations while

The slogan 'success breeds success', and the idea that the more a document is cited the more it will get cited, are not true; there is generally a steep decline in citedness

the average paper only receives 0.3. The slogan that 'success breeds success' and the idea that the more a document is cited, the more it will get cited are, however, not true. Even for previously highly cited papers, there is generally a steep decline in citedness five years after publication.

Differences in citation life-cycle curves

To what extent do the citation life-cycle curves of the individual highly cited papers differ? In order to analyse our data to answer this question we classified the articles in different clusters according to their citation patterns. To adjust for the large differences in the citedness of the individual papers, we used relative shares and not absolute numbers as the basis for our analysis. First we identified the citations received annually using a 12-year citation window. In order to obtain this long-term citation window, only the 1981–1989 publications were included in the analysis (137 papers). We then calculated the percentage of the citations received in two periods: one to three years after publication, and seven to 12 years after publication. In the next step we classified the articles in different categories. If more than 30% of the citations were obtained during the first three years, the paper was classified as an 'early rise' paper. Similarly, if less than 15% of the citations were obtained in this period it was classified as a 'delayed rise' paper. Intermediate papers (percentages from 15 to 30) were classified as 'medium rise' papers. In the latter period a percentage of less than 30 corresponded to a 'rapid decline' paper. If more than 50% of the citations were obtained during these last six years it was termed a 'no decline' paper. Finally, a percentage from 30 to 50 corresponded to a 'slow decline' paper. The distribution of the highly cited articles is shown in Table 3.

As we can see, two categories account for the large majority of the papers: One type of papers is characterised by a moderate period of initial increase followed by a gradual decline (medium rise – slow decline). The other type is characterised by a relatively slow rise, with a stable citation level thereafter (delayed rise – no decline). Also, the categories 'early rise – rapid decline' and 'medium rise – no decline' account for a significant share of the papers. The other categories are non-existent or marginal.

Table 3. Percentage of highly cited papers in different citation life cycle clusters*

	Rapid decline (<30)	Slow decline (30–50)	No decline (>50)
Early rise (>30)	11%	4%	0%
Medium rise (15–30)	3%	37%	10%
Delayed rise (<15)	0%	4%	31%

Note: * Based on the 1981–1989 publications, and a 12- year citation window

The citation curves of the three largest clusters have been showed in Figure 4 (plotting average values for each year).

It appeared that the majority of the least cited papers represented 'early rise – rapid decline' papers, while most of the papers that in absolute terms obtained the highest number of citations during the 12-year period represented 'delayed rise – no decline' papers (which is not unexpected since a high citation rate during the entire period is a prerequisite for obtaining a very large number of citations).

Former studies of citation life cycles have reported results resembling ours (Aversa, 1985; Cano and Lind, 1994). For example, in a study of 400 highly cited papers Aversa found that a delayed rise in citedness is associated with less rapid ageing and a higher number of total citations, and that an early rise in citedness is associated with fewer total citations and a very rapid ageing rate (see also Arunachalam and Singh, 1984; Line, 1984). However, Aversa identified only two basic citation patterns: delayed rise – slow decline and early rise – rapid decline. Also, according to Aversa's findings few papers showed no evidence of ageing, in contrast to our results in which 41% (31% + 10% cf. Table 3, right-hand column) of the papers showed no or only a little decline in citedness during a 12-year period.

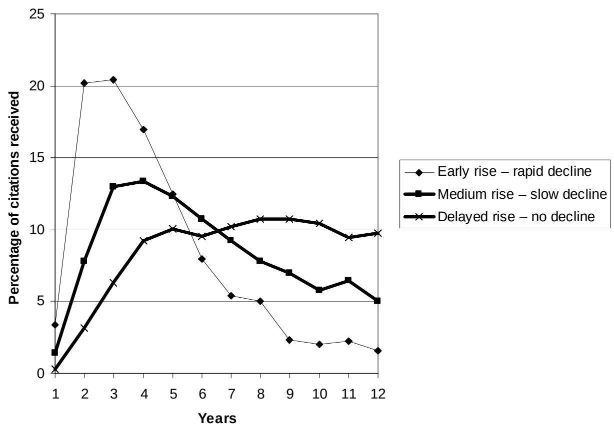


Figure 4. Citation curves for three different clusters of highly cited papers*

Note: * Based on the 1981–1989 publications, and a 12-year citation window

We also analysed the distribution of the articles by subject areas. The distribution is shown in Table 4, using ISI's Current Contents categories. As we can see, the distribution shows variations among the subject areas. The difference is largest between biology and the life sciences on the one hand, and the physical, chemical and earth sciences on the other. In the first fields the 'no decline' papers dominate and no papers show a rapidly declining citation pattern. In the latter fields, 40% of the papers are rapidly declining. Possibly, the rapid decline papers are within areas in which the research front is fast-moving. In contrast, the no decline papers are likely to present research of continuing interest (e.g. by paradigm articulation or by developing particularly useful methods) so that the key publications would be referred to again and again. This can be no more than a first-round suggestion, because life sciences are usually taken to be fast-moving fields.

In conclusion, the citation curves of highly cited papers are not uniform. The differences between the citation curves of individual papers are being concealed in aggregated patterns. Also within the different clusters there are large individual differences among the publications. These differences may be caused by intrinsic differences related to the nature of the papers and their cognitive functions, and may be important variables in the explanation of frequent citations.

Citations from adjacent and remote fields

The ability of a paper to be cited by adjacent fields as well as its own field has been argued to be an important factor in explaining frequent citations (Seglen, 1997b). Knowledge on this issue, however, appears to be limited. We were interested in to what extent the highly cited Norwegian articles were cited by remote fields.

In order to answer this question we made a number of different analyses of the citing papers, that is, the papers that have cited the highly cited articles. Using bibliometric information we analysed from which journals the citations were received and the field assignments of these journals. First we collected citation data counting all citations from year of publication through 2000. Altogether the 297 articles collected 51,078 citations. For each article we counted the number of different journals citations were received from and the total number of citations.

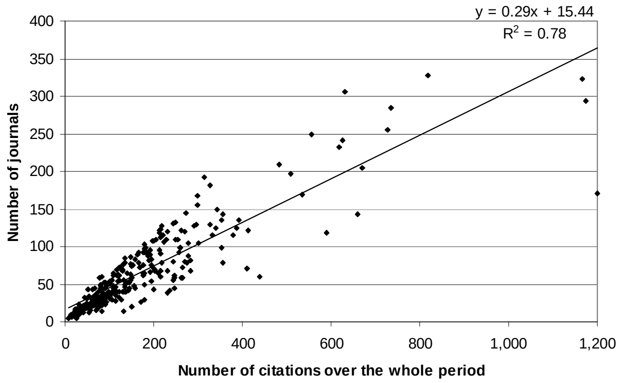


Figure 5. Relationship between number of citations and number of citing journals for each of the highly cited papers

Generally, we find that the highly cited papers were obtaining citations from a large number of journals; on average each paper was cited by 65 different journals. However, there were large differences among the highly cited papers. Not unexpectedly, we find a correlation between the total number of citations and the number of different journals (R^2 = 0.78, Figure 5). Still there were individual variations. For example, the papers that have received between 600 and 700 citations were cited by 143 to 306 different journals. For almost all papers, the journal citation distribution is very skewed. There are some central journals contributing a large share of the citations and many peripheral journals contributing only one or a few citations. Furthermore, it should be added that the range of journals

Table 4. Percentage of highly cited papers in different citation life cycle clusters by subject areas*

	Early rise – rapid decline	Early rise – slow decline	Medium rise – rapid decline	Medium rise – slow decline	Medium rise – no decline	Delayed rise – slow decline	Delayed rise – no decline
Biology and enviromental sciences, agriculture		4%		42%	15%		38%
Clinical medicine	15%	3%	3%	42%	9%	7%	20%
Life sciences				53%	13%		33%
Physical, chemical and earth sciences	33%		7%	26%	7%		26%
Engineering, computing and technology	11%	7%	4%	25%	7%		46%
Average	**11%**	**4%**	**3%**	**37%**	**10%**	**4%**	**31%**

Note: * Based on the 1981–1989 publications, and a 12-year citation window probably becomes larger over time (which one does not see at the aggregated level). This is because of visibility effects. When a paper gets highly cited it becomes visible outside its original research field, and attracts attention and citations from an ever broader range of fields.

In Table 5 we have classified the citations according to the 'distance' of the fields they represent. Field 'distances' have been calculated using percentage of citations as a measure. Fields that in the national subset of articles contribute to more than 30% of the citations in a field have been termed 'close fields'. Similarly, we have calculated the distribution in other categories: 'related fields', 'intermediate fields', 'remote fields' and 'very remote fields' (see notes below Table 5 for details). We have also calculated the distributions of a comparable sample of 'average' papers (i.e. for the overall national publication set, weighted according to the field distribution of the highly cited papers).

The results show that the highly cited papers have an almost identical citation distribution to the average sample. It may seem surprising that the highly cited papers do not get relatively more citations from remote fields. Still, more than 40% of the citations to the highly cited papers come from remote or very remote fields — but this is also the case for the average sample. This means that the pattern of contributions close/remote is determined by the nature of the field, rather than being highly cited or not.

At the level of the individual articles, however, the picture might look different. For example, one highly cited paper may be cited by close as well as remote fields. Such a citation pattern is probably unusual for an 'ordinary' cited paper. One hypothesis would be that some of these papers are cited by close fields only, other papers are cited by remote fields only (because their contents are outside the field they have been assigned to). In sum this would result in an overall citation pattern similar to the one found for the highly cited papers.

Perhaps the most urgent question arising from Table 5 is why papers get so many citations from remote fields. According to the above considerations it can be explained by internal differences in the publication set and the effect caused by papers being outside the core-areas of the fields. Knowing that ISI's subject classification system has major weaknesses (Aksnes *et al*, 2000), this seems likely. Furthermore, other studies of knowledge transfer have shown the classification system being applied to be influential: In a study of Porter and Chubin (1985), citations crossing broad field categories were found to be extremely infrequent. Using a less broad classification system, Rinia *et al* (2002) found a much higher incidence of cross-disciplinary citations. In our study, we used a rather fine-grained system of categories. In consequence, a much lower percentage would appear as 'remote' compared with only counting citations crossing broad field categories.

A related question concerns the content of the highly cited articles. Probably, in order to get highly cited the content of the highly cited paper must be useful or of relevance for the research activity in many research fields. This means that articles bursting the limit of their own field can obtain a very large number of citations. By contrast, papers addressing very specialised issues that are being covered by only a few journals will not obtain a large number of citations. In this way, the size of the field and its annual scientific production is also influential. Examining the journal list of each paper we find it difficult to make any reliable conclusion on the content of the articles. Content analyses of the individual papers would be necessary in order to provide more knowledge on this issue.

Discussion

In our study we have examined the bibliometric characteristics of articles by Norwegian researchers that are highly cited within their fields. Although limited in scope, the papers analysed represent the very top of a publication set of more than 45,000 publications. Various hypotheses concerning highly cited papers have been examined. We have found that these papers typically are authored by a large number of scientists, often involving international collaboration. The citation curves of the papers generally follow a typical pattern of rise and decline, with a steep decline in citedness five years after publication. Nevertheless, there are large differences in the citation pattern of the individual papers. Although most of the papers are published in high-impact journals, there are also papers present in

Table 5. The distribution of citations by field distances, highly cited papers vs. average sample, Norway 1981–1996

	Close fields	Related fields	Intermediate fields	Remote fields	Very remote fields
Highly cited papers	24.9%	14.4%	20.7%	31.5%	9.1%
Average sample*	25.5%	14.4%	20.3%	29.7%	10.7%

Notes: * Weighted according to the average for an identical sample of papers (similar field distributions)

Field 'distances' have been calculated using percentage of citations as measure. Fields that contribute to more than 30% of the citations in a field have been termed 'close fields'. For the other categories the percentages are: related fields: 15–30%, intermediate fields: 5–15%, remote fields: 1–5%, very remote fields: <1%. As the basis for these calculations we have used the citation patterns of the overall national publication set.

High citation scores are the results of many researchers' decisions to cite a particular paper; quality and visibility dynamics can be useful in order to understand some of the mechanisms involved

poorly cited journals. The large majority of the highly cited papers represent regular journal articles, although the share of review articles is higher than in the overall national subset.

On several issues our results have confirmed findings of former studies. This suggests that highly cited papers have certain recurrent common characteristics. Still, the question of generality remains uncertain. First our study is based on a particular definition of highly cited. Another definition or set of criteria would give a different sample and thereby partly change the characteristics identified. Second our analysis is carried out for one nation only — Norway (although two-thirds of the highly cited articles were co-authored by foreign scientists). Norway is well integrated into the international scientific arena, and Norwegian researchers collaborate extensively with researchers abroad. As a scientific nation Norway is rather small and is among the less R&D-intensive countries in the OECD. The country has research activities in a broad range of scientific specialities, but is specialising in research related to its natural resources. Because of the high share of international co-authorship among the highly cited papers and the relatively low percentage of citations from Norwegian authors (also for all papers), we still think the peculiarities of Norway as a research nation should not be given too much emphasis when interpreting the results.

In our study there are several intriguing findings, which allows a further step: explanation of patterns in highly cited papers. In the last part we will discuss how these findings can be explained by introducing a conceptual distinction between quality dynamics and visibility dynamics.

Quality and visibility dynamics

High citation scores are the results of many researchers' decisions to cite a particular paper. In order to explain how highly cited papers come about, one has to focus on how such micro-level processes aggregate. The concepts of quality and visibility dynamics can be useful in order to understand some of the mechanisms involved.

The quality dynamic is related to the structure of scientific knowledge. Generally, scientific progress is powered by a variety of contributions. Some represent major scientific advances; others are filling in the details. A fruitful distinction in order to analyse this structure is the one between *core knowledge* and *frontier knowledge* (Cole, 1992). According to Cole, core knowledge consists of the basic theories within a field, while frontier knowledge is knowledge currently being produced. A large part of what is published does not as such pass its way into core knowledge. This is because much of the research produced at the frontier aims at producing low-level descriptive analyses or represents contributions that turn out to be of little or no lasting significance (Cole, 2000). Similarly, much of what is published represents 'dead ends' and does not function as a basis for further knowledge development (although in order to make progress, you may have to explore or know about these matters). In consequence, one expects a skewed distribution of citation scores, and differences between fields depending on the relationship between evolving core knowledge and more ephemeral frontier knowledge.

Citation is a social process. Generally, there may be a large number of reasons for why an author has cited a particular paper (Cronin, 1984). A concept of cognitive differentiation is still relevant because scientists (in a very simplistic model) tend to cite contributions that are useful for their own research. Accordingly, papers receiving many citations have been useful for many scientists (the quality dynamic is socially determined). Citations are, however, not a simple reflection of this quality dynamic. This is because the citation process is influenced by a variety of other mechanisms. By visibility dynamics we mean certain social mechanisms that influence the citation rates. Particularly important is the bandwagon effect. When one paper is cited by many subsequent papers, even more people become aware of this paper. Thus, its visibility increases and thereby the chances of getting even more citations. This is a variant of the 'Matthew effect' (Merton, 1968), stating that recognition is skewed in favour of established scientists. Similarly, when a paper has received many citations the paper obtains status as an authoritative paper. In turn even more scientists will cite it, since appealing to existing authorities may be one reason for citing a paper (Gilbert, 1977).

How can our findings be explained?

Some of our findings can (partly) be explained by using the concept of quality dynamics. For example, if highly cited papers often involve collaboration between a large number of researchers, this is because larger groups have more human resources available and can produce scientifically important results (see e.g. Melin, 1997). Similarly, international collaboration will often contribute to an increase in the scientific power of the groups, because of the additional economic, human or technical resources accessible. In particular, needs for costsharing and access to facilities abroad may be of importance for small countries like Norway. Furthermore, quality dynamics can partly explain the different citation life-cycles of the highly cited papers. For example, some papers are initially important but are rapidly superseded by later works. These are the 'rapid decline' papers. Other papers are important and of continuing interest in a gradually evolving core. These are the 'no decline' papers.

Visibility dynamics, on the other hand, are relevant for the explanation of other results of our study. For example, papers authored by many scientists have higher visibility than single-author publications (through the personal communication of many authors), which increases the possibility of obtaining more citations — although, as we have seen, another explanatory factor is the mechanism of the quality dynamics.

Visibility dynamics also affect the citation life-cycles of highly cited papers. As described above, these dynamics imply a bandwagon effect in terms of citation. This process can partly explain why the papers generally are frequently cited many years after publication. Still, it should be added that after 10 years the highly cited papers obtain only half as many citations annually as they do in their peak years. The meaning of this decline depends on the field and its rate of obsolescence. What is cognitive ageing in one field, and thus does not reflect negatively on quality, will be an indicator that Matthew effects were at play in another field.

With respect to the different life-cycle patterns identified, visibility dynamics are probably most influential for the 'no decline' papers. This is because the mechanisms related to the visibility of a paper operate most strongly after some years when the paper has obtained an authoritative status. Being authoritative is thus a matter of ascription, rather than an inherent quality of the paper. But having authoritative papers to refer to in a field is important for its functioning (coherence), even if it is not a simple measure of quality. To the contrary, the rapid decline papers are not likely to benefit from such effects.

Also, visibility mechanisms are relevant in the explanation of the journal distribution of the highly cited articles. Generally, articles published in journals with high-impact factors are likely to obtain higher visibility than articles published in less cited journals. As we have seen, most of the highly cited articles are published in high-impact journals. However, even within high-impact journals there is usually large variability in the citation rates of the articles, most of the publications being poorly cited (Seglen, 1992). Thus, although visibility is higher in a high-impact journal this cannot *per se* explain high citation counts — quality dynamics are also important.

Interestingly, Figure 5 can be interpreted as showing both dynamics. For the papers above the regression line, the number of journals in which they are cited increases faster than the total number of citations. The citations are spread more thinly than the regression line predicts, which has to do with a larger variety of authors wanting to profit from the visibility of the highly cited paper. Below the regression line, the reverse is true. There is more in-depth citing, which one can link with quality dynamics (even if some visibility dynamics will always be present as well).

These examples show how several of our findings can be interpreted by using this conceptual distinction. Still, the conceptual model has obvious limitations. For example, an important or useful paper in a small, narrow or highly specialised field will receive much fewer citations than a similar paper in a more popular field — although this can be considered as a problem of working with standardized indicators, not of the model. A more sophisticated model would, however, have to be based on a more complex analysis of the citation process.

Concluding remarks

In conclusion we find a complex pattern in which both quality dynamics and visibility dynamics contribute to the citation counts of highly cited papers. The empirical findings of this study do not allow any definite assessment on the relative importance of the two dynamics. As we have seen there are important variabilities at individual levels. Still, we would like to emphasis that mechanisms such as the bandwagon effect increase the variability in citation rates between scientific contributions. Some contributions will benefit greatly from this effect while others will not. In turn, the skewness in the citation distribution is larger than the quality differentiation among scientific contributions might justify. To what extent this is the case is, however, an issue that needs further investigation.

With these conclusions we have identified a challenge for quality assessment of scientific publications based on citation counts. There are good reasons to do this, but quality dynamics and visibility dynamics are inextricably linked. Simple corrections like deleting self-citation are not enough (and sometimes misguided). What one can do and should do is to base such assessments on a study of citations, rather than on indicators. This goes against the need of working efficiently, but what is the use of being efficient in the production of indicators that have little value? This paper has shown that there are interesting patterns, and that indications (rather than indicators) of research quality can be given when one interprets citation counts against such patterns.

Notes

1. Through the available data it has been possible to test this assumption for one single year: 1996. The selected sample includes 20 articles from 1996. When instead using a standard five-year citation window as basis for the calculation,

we identified exactly the same 20 articles on the top of the list. Thus, this indicates that using a running instead of a standard five-year citation window gives an identical (or almost identical) sample. For our purpose the method of selection must therefore be regarded as satisfactory.

2. The strong element of collaboration in the sample of highly cited papers is also evident from bibliometric data on number of addresses. These data show that in 18% of the articles one single address was indexed. These papers are therefore likely to represent single-institute productions. This percentage is much higher in the total population of papers (42%).
3. For example, for an article published in 1981 we used the impact factor for the journal the article appeared in for the period 1981–1985, calculated as the inverse ratio between the number of articles (counting: articles, notes, reviews, and proceeding papers) and the number of citations to those papers within the same five-year window. We then collected the corresponding average citation rate for the field the journal was assigned to and calculated an index number.

References

D W Aksnes (2003), 'A macro study of self-citation', *Scientometrics*, 56(2), pages 235–246.

D W Aksnes, T B Olsen and P O Seglen (2000), 'Validation of bibliometric indicators in the field of microbiology: a Norwegian case study', *Scientometrics*, 49(1), pages 7–22.

D W Aksnes and G Sivertsen (2004), 'The effect of highly cited papers on national citation indicators', *Scientometrics* (forthcoming).

S Arunachalam and U N Singh (1984), 'Publication and citation patterns in the literature of high metabolism area: the case of superconductivity in 1970', *Journal of Information Science*, 8, pages 93–102.

E S Aversa, E S (1985), 'Citation patterns of highly cited papers and their relationship to literature aging: a study of the working literature', *Scientometrics*, 7(3–6), pages 383–389.

V Cano and N C Lind (1994), 'Citation life cycles of ten citation classics', *Scientometrics*, 22(2), pages 297–312.

S Cole (1992), *Making Science: Between Nature and Society* (London: Harvard University Press).

S Cole (2000), 'The role of journals in the growth of scientific knowledge', in B Cronin and H B Atkins (editors), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (Medford: ASIS).

B Cronin (1984), *The Citation Process: The Role and Significance of Citations in Scientific Communication* (London: Taylor Graham).

European Commission (2001), *Key Figures 2001. Special edition. Indicators for benchmarking of national research policies* (Brussels).

N G Gilbert (1977), 'Referencing as persuasion', *Social Studies of Science*, 7, pages 113–22.

W Glänzel and H-J Czerwon (1992), 'What are highly cited publications? A method applied to German scientific papers, 1980–1989', *Research Evaluation*, 2(3), pages 135–141.

W Glänzel, E J Rinia and M G M Brocken (1995), 'A bibliometric study of highly cited European physics papers in the 80s', *Research Evaluation*, 5(2), pages 113–122.

S M Lawani (1986), 'Some bibliometric correlates of quality in scientific research', *Scientometrics*, 9(1–2), pages 13–25.

M B Line (1984), 'Citation decay of scientific papers: variations according to citations received', *Journal of Information Science*, 9, pages 90–91.

M H MacRoberts and B R MacRoberts (1989), 'Problems of citation analysis: a critical review', *Journal of the American Society for Information Science*, 40(5), pages 342–349.

B R Martin and J Irvine (1983), 'Assessing basic research: Some partial indicators of scientific progress in radio astronomy', *Research Policy*, 12, pages 61–90.

G Melin (1997), *Co-Production of Scientific Knowledge: Research Collaboration between Countries, Universities and Individuals* (Umeå: Umeå University).

R K Merton (1968), 'The Matthew effect in science', *Science*, 159, pages 56–63.

H P F Peters and A F J van Raan (1994), 'On determinants of citation scores: a case study in chemical engineering', *Journal of the American Society for Information Science*, 45(1), pages 39–49.

A L Porter and D E Chubin (1985), 'An indicator of cross-disciplinary research', *Scientometrics*, 8(3–4), pages 161–176.

D d S Price (1965), 'Networks of scientific papers', *Science*, 149, pages 510–515.

D d S Price (1976), 'A general theory of bibliometric and other cumulative advantage processes', *Journal of the American Society for Information Science*, 27, pages 292–306.

E J Rinia, T N van Leeuwen, E E W Bruins, H G van Vuren and A F J van Raan (2002), 'Measuring knowledge transfer between fields of science', *Scientometrics*, 54(3), pages 347–362.

P O Seglen (1992), 'The skewness of science', *Journal of the American Society for Information Science*, 43(9), pages 628–638.

P O Seglen (1997a), 'Citations and journal impact factors: questionable indicators of research quality', *Allergy*, 52(11), pages 1050–1056.

P O Seglen (1997b), 'Why the impact factor of journals should not be used for evaluating research', *British Medical Journal*, 314(7079), pages 498–502.

R J W Tijssen, M S Visser and T N van Leeuwen (2002), 'Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference?', *Scientometrics*, 54(3), pages 381–397.

A F J van Raan (2000), 'The Pandora's box of citation analysis: measuring scientific excellence — the last evil?', in B Cronin and H B Atkins (editors), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (Medford: ASIS).

5. ARTICLE III. A macro study of self-citation.

Aksnes, D. W.

Published in <u>Scientometrics</u> 56(2) (2003): 235-246.

Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 56, No. 2 (2003) 235–246

A macro study of self-citation

DAG W. AKSNES

Norwegian Institute for Studies in Research and Higher Education (NIFU), Oslo (Norway)

This study investigates the role of self-citation in the scientific production of Norway (1981-1996). More than 45,000 publications have been analysed. Using a three-year citation window we find that 36% of all citations represent author self-citations. However, this percentage is decreasing when citations are traced for longer periods. We find the highest share of self-citation among the least cited papers. There is a strong positive correlation between the number of self-citations and the number of authors of the publications. Still, only a minor part of the overall increase in citation rates that can be found for multi-authored papers is due to self-citations. Also, the share of self-citation shows significant variations among different scientific disciplines. The results are relevant for the discussion concerning use of citation indicators in research assessments.

Introduction

Self-citations account for a relatively large share of all citations. Given the cumulative nature of individual research, citing oneself may be considered as a natural and acceptable procedure.[1] On the other hand, scientists also tend to cite themselves as a result of egotism, for establishing their own scientific authority or to make their former works visible.[2] Particularly when citations are used as indicators for assessing scientific impact, these citations are often treated as problematic.[3,4] The reason is that self-citations do not reveal much about the impact of a work in the scientific community. Against this background it is important to know how prevalent self-citation is and how it influence on the citation indicators.

A self-citation is usually defined as a citation in which the citing and the cited paper have at least one author in common. However, the term is sometimes used for other kinds of citation linkages (e.g., journal self-citations and institutional self-citations[5]). Also, in a more restricted version only publications having identical first authors are included as author self-citations.[6] Methodologically, self-citations can be counted in two different ways, diachronously or synchronously.[2] An author's synchronous self-citations are those contained in the citations the author gives, that is in the reference lists of his publications. The diachronous self-citations are those included in the citations an

Received August 12, 2002.

Address for correspondence:
DAG W. AKSNES
Norwegian Institute for Studies in Research and Higher Education (NIFU),
Hegdehaugsveien 31, NO-0352 Oslo, Norway
E-mail: Dag.W.Aksnes@nifu.no

0138–9130/2003/US $ 20.00

author receives. While an author's synchronous self-citation rate can be calculated by considering the papers he has authored or co-authored, a citation database must be adopted in order to estimate diachronous self-citation rates.

Several studies have analysed self-citations synchronously. According to *Tagliacozzo*[7] the share of self-citations amounted to approx. 17% in the two fields plant physiology and neurobiology. In another synchronous case study, *Lawani*[2] found a self-citation rate of 15% in agronomic literature, while the corresponding share in cancer literature was 10%. *Bonzi* and *Snyder*,[8] in a case study including the social sciences and the humanities, found an average self-citation rate of 11%, varying from 16% in the physical sciences (chemistry and geology) to 3% in the social sciences (economics and sociology). Rather similar results were presented by *Snyder* and *Bonzi*.[9]

These studies have shown that self-citation rates tend to vary between disciplines. A major drawback with all the studies, however, is their small sample size. For example, *Tagliacozzo*'s study[7] involved approx. 180 articles, while *Bonzi* and *Snyder*[8] analysed only 120 publications. Still, it appears that a synchronous self-citation rate of 10 to 20% is typical within the natural sciences and medicine.

Diachronous self-citation rates, on the other hand, may differ from those calculated synchronously. In order to estimate such self-citations rates one needs bibliometric information on the citing papers. In some bibliometric studies of research performance the percentage of diachronous self-citations has been included as an indicator. For example, a case study of physics in the Netherlands (1985-1994) found a self-citation rate of 29%.[10] Similarly, 29% of the citations in a study of Dutch chemistry (1980-1991) represented self-citations.[11] The percentage of self-citations has also been shown to be highest the first years after publication.[12] Still there is need for more systematic information on self-citation rates. This is particularly important for assessing the representativity of citation indicators as performance measures.

The aim of this study is to contribute to the knowledge on self-citation, and self-citation rates in particular. In the study we analyse the scientific production of Norway (1981-1996), in total more than 45,000 publications. We assess the element of self-citation in this scientific production and how the share of self-citation varies according to different parameters such as scientific discipline, overall citedness and number of authors. Finally, we discuss the implications of the findings with respect to the use of citations as indicators.

Methods

As basis for our analyses we used the ISI-database National Citation Report (NCR) for Norway. This database contains bibliometric information for Norwegian articles (papers with at least one Norwegian address). In addition the database includes data on the citing papers, that is, the papers that have cited the Norwegian publications. We identified the Norwegian subset of articles included in the ISI-product National Science Indicators (NSI) (counting articles, notes, reviews, and proceeding papers). Articles within the social sciences and the humanities, however, were excluded from our analysis. We applied the 2001 editions of the ISI-products with data covering 1981-2000. We extracted the publications from the 15-year period 1981-1996 and identified the citations to these articles from year of publication through 2000. In this way the citation window applied varied from 5 to 20 years, with an average citation period of 11.6 years. For each article we calculated the number of citing papers that represented author self-citations, using the criterion that at least one author (first author or co-author) is also an author (first author or co-author) of the citing paper. On this basis a variety of different bibliometric analyses were carried out.

Results

In total 46,849 articles were analysed. 4,845 articles were uncited. Of the remaining 42,004 articles 71% (29,842) had one or more self-citations. From year of publication through 2000 the Norwegian subset of articles received a total of 640,710 citations (average 13.7 citations per article), of these 136,316 were author self-citations. Thus, in the national subset of articles the self-citation share amounted to 21%.

Poorly cited papers have the highest self-citation shares

In the first study we analysed how self-citation varies with the overall citation rates of the publications. All the 42,004 cited articles were included in this analysis and they were divided into groups according to their overall citation rates. For each group we calculated the average numbers and shares of self-citations. The results are shown in Figure 1. For better visibility only publications with 50 or less citations have been included in the figure (39,957 papers). As we can see the number of self-citations increases with total number of citations. For instance, papers cited 5 times on average received 1.4 self-citations. In contrast, publications cited 50 times received 11 self-citations.

In relative terms, however, we find the highest share of self-citations among the least cited papers. For example, this share is 29.9% for papers cited less than 5 times, while it is 19.4% for papers cited 46-50 times. The variables are strongly correlated with an R-square value of 0.88 (using the publications cited 1-50 times as basis for the calculation). For papers cited more than 50 times, the self-citation rate is even lower (14.5%). (In the figure we can see an upswing at 50 times cited, but this is not a continual trend.)

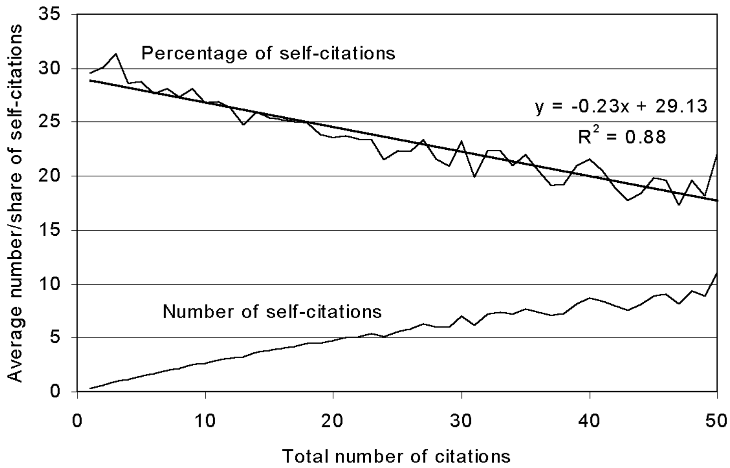


Figure 1. Self-citations (average number and share) as a function of article citedness (n=39,957)

It is not unexpected that the share of self-citation is decreasing with number of citations. There are practical limits (frequency of publication) for how often the author(s) can possibly cite one publication. In fact, it is interesting that the differences in self-citation shares are not larger. Intuitively one could expect an even higher percentage of self-citation for publications with very few citations, and an even lower percentage for the highly cited papers. For the set of publications cited only once, it would not be surprising if the majority of these citations represented self-citations (i.e., if the poorly cited papers have little scientific importance, mainly the authors would want to cite them). Our results, therefore, are interesting by showing that this is not the case.

Still, as noted above, 29% of the cited articles did not receive a single self-citation. The large majority of these "non-self-cited" articles are poorly cited. In the overall publication set (when uncited articles also are included) 37% of the articles did not receive any self-citation. This share may seem surprisingly high. It means that the self-citations are essentially unevenly distributed in the population of articles. On the other hand, it is likely that the share of articles not containing any self-citation would be lower if citations were counted synchronously. That is, it would be more unusual if an author does not include any reference to his former works when he writes an article. This follows from the fact that authors tend to cite some of their publications often, others not. In turn there are different reasons involved. When an article does not contain any self-citation (or more precisely a self-reference) the reason may be that the author(s) has no relevant former works within the area. Alternatively, it can be related to the content of the article, for example if a paper is only presenting new data. When an article has not received any self-citation the reason may be that the author did not continue the research in the particular area or more generally, that the article did not represent a contribution of relevance in his further research activities.

Multi-authored papers receive more self-citations

In another study we analysed how the number of self-citations varied according to the number of authors of the publications. Here, one would expect that the larger the number of authors, the larger the number of self-citations (more authors to cite themselves). In fact, such a correlation has been demonstrated in a few former studies.[8,9] Also in our study we found the number of self-citation to increase with number of authors (Figure 2). For example, while single-author papers on average received 1.5 self-citations, papers with 10 authors received 6.7.

It is generally known that the total number of citations also increases with number of authors.[13] To what extent can this increase be explained by self-citations? In order to assess this question we calculated the overall citation rates for the publications (Figure 2). As we can see, the citation rates increase rapidly with number of authors. For each additional author the papers on average receive 2 more citations. However, the increasing number of self-citations can only explain a minor part of this general increase in citation rates.

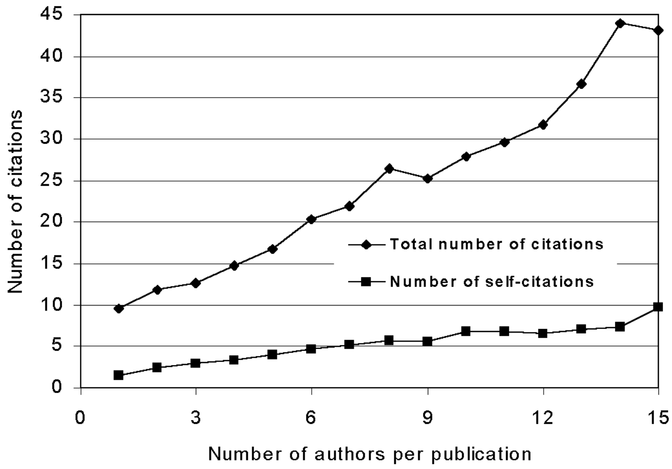


Figure 2. Number of self-citations and total number of citations as a function of number of authors per paper (n= 46,425*)
* 424 articles with more than 15 authors have not been included in the figure

Using citations as measures of scientific impact, our study in this way indicates that it "pays to collaborate". The results contradict the conclusion by *Herbertz*[13] that differences in citation rates with regard to cooperation can be explained by self-citations. However, his study was of a limited scope, involving another methodology. Also, *van Raan*[14] has questioned *Herbertz*'s conclusions.

The share of self-citation varies among scientific disciplines

We also analysed how the share of self-citation varied according to scientific fields. For each publication we collected data on field assignments. As basis for the classification we used the NSI-categories (Standard version).

Overall, we find considerable variations in the self-citation rate among different scientific fields (Table 1). The lowest percentage is in Clinical Medicine (17%), while the highest percentage is in Chemistry and in Astrophysics (31%). There are no obvious explanations for these variations. No correlation could be found between the share of self-citation and the average citation rate of the fields (left column) (R^2=0.02). Similarly, the self-citation share did not correlate with the average number of authors per publication (R^2=0.08). On the basis of our other findings these results may seem surprising. Accordingly, at field levels there are other more important factors that influence on the self-citation rates. For example, field variations in citation norms may also affect the tendency to cite oneself. Furthermore, the extent of cumulative work in various fields can possibly explain the difference. More cumulativeness at the level of individual research, are likely to cause more self-citation. As we have seen, the highest percentages are found in the "hard sciences". Here explanation, and thus building on earlier work are important, in contrast to disciplines with a larger element of descriptive research (data collection etc.). Finally, at subfield levels one would probably find that scientists working in narrow specialities have the highest self-citation counts.[15]

Table 1. The percentage of self-citations in different scientific fields (using NSI-categories).

Field (NSI-category)	Citation rate	% Self-cit.	No. of papers
Clinical Medicine	14.2	17 %	15,909
Neuroscience	22.6	18 %	1,934
Ecology/Environment	14.3	19 %	1,835
Plant & Animal Science	12.0	20 %	5,040
Psychology/Psychiatry	9.8	21 %	1,325
Geosciences	12.4	21 %	2,898
Microbiology	18.3	21 %	1,246
Materials Science	6.2	22 %	774
Agricultural Sciences	8.8	22 %	1,051
Mathematics	7.9	22 %	697
Molec Biology & Genetics	24.8	22 %	1,132
Biology & Biochemistry	19.5	23 %	4,153
Engineering	6.5	23 %	2,072
Pharmacology	11.1	23 %	1,390
Multidisciplinary	10.6	24 %	572
Computer Science	4.9	24 %	283
Immunology	22.0	24 %	1,579
Physics	10.9	26 %	2,857
Chemistry	11.0	31 %	4,530
Astrophysics	15.6	31 %	383

Many self-citations among first year citations

In the last study we analysed how the number and share self-citations varied with time after publication. For each article we identified the self-citations by citing years, that is, by year after the article was published. We then calculated the percent of self-citation for each citing year. The results of the analysis are shown in Figure 3.

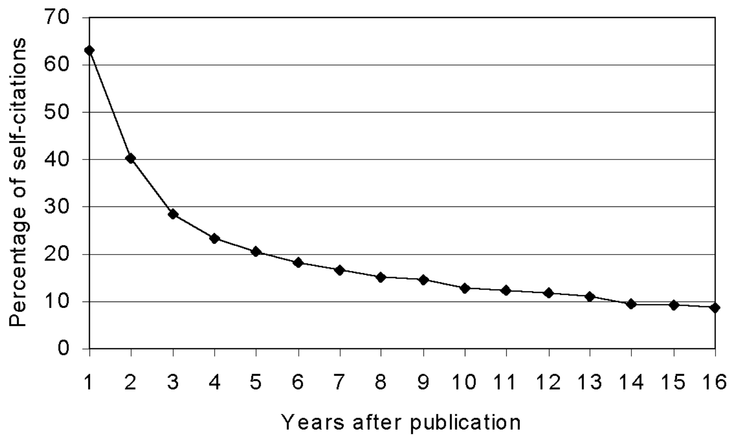


Figure 3. Percent of self-citation by year after publication (n=46,849*)
* For the 1-5 year period all articles (46,849) have been included in the analysis. For the subsequent years the number of articles have been reduced according to the citation period the articles have obtained

As is evident the distribution is skewed. The majority of the citations the articles obtain the same year as publication are self-citations (63%). Also among the second-year citations there are many self-citations (40%). In contrast 15 years after publication only 9% of the citations represent self-citations.

However, in absolute terms the picture is changing. For example, the same year as publication, 7,900 of 12,600 citations represent self-citations. The number of self-citations increases to 22,100 in the second year and reaches its peak the third year after publication (23,100). In comparison, the total number of citations is highest in the fourth year.

Often in bibliometric analyses, citations are collected during the first three years after publication. In our national publication set this would result in a self-citation share of 36%. In contrast, the self-citation share would be 29% when using a 5-year window.

These findings have important implications. When a short-term citation window is used, self-citation will be a larger problem than when using a long-term window. For example, the widely used Journal Impact Factor (calculated as the mean number of citations in a given year, to journal items published during the preceding 2 years) will be highly influenced by self-citations.

The results are also relevant for the interpretation of our other findings. For example, the percentages of self-citation in different scientific fields would be higher if choosing a shorter citation window. In conclusion, it is important to appreciate that short-term impact as measured by citations is heavily influenced by self-citations.

Discussion

Self-citation may be regarded as an ambiguous phenomenon. On the one hand citing oneself is natural and may have important functions, for example by connection the present work of an author to his/her former works within the area. A study of the motives for citation showed that there were virtually no differences between the reasons that authors cite their own work and the reasons they cite other works.[16] On the other hand, self-citation may also be a reflection of egotism. For example, authors may tend to gratuitously cite their own works in order to raise their citation counts or to make their former works visible – although there are practical (frequency of publication) as well as normative limits for how often one can cite oneself. *Garfield* and *Welljams-Dorof*[17] have accordingly emphasised that excessive self-citation are readily apparent to the readers and should be corrected in the editorial and peer review process.

Generally, the self-citation shares found in this study are higher than the results of former synchronous case-studies.[2,7–9] Thus, self-citation appears to be more widespread than these studies have indicated. However, the basis for comparison differs since we have measured self-citations diachronously. Theoretically, synchronous and diachronous methods should give identical results within the same set of publications. But in practice the methodological basis will differ. For example, the choice of a particular citation window influences on the self-citation rates in diachronous studies, as we have seen.

However, it should be emphasised that our analysis is carried out for one nation only: Norway. The results obtained may, therefore, not have general validity. Further investigations are required in order to assess this question. Still, representing a large-scale analysis involving more than 45,000 publications the overall patterns identified are likely to be typical. We think it is unlikely that Norwegian researchers differ radically in citation behaviour from other researchers. In particular, this is due to the fact that Norwegian science has a strong international orientation: 29% of the publications analysed have authors also from other countries, and this share has been increasing during the entire period.

In a science policy context the question is to what extent self-citation represents a problem. When citations are used as indicators, these citations may be considered problematic, since they hardly can be considered to reflect any impact of a work in the scientific community. In fact, when our results show that 20 to 35 percent of all citations typically are self-citations this may undermine some of the basis for using citations as indicators of scientific impact. As we have seen this has also been an argument from critics of citation indicators.

Still self-citation may not be a serious problem at aggregated levels. For example, *Phelan*[1] has argued: "It is probably the case that self-citation is not a major problem at most level of analysis. For this phenomenon to influence conclusions in studies of universities or nations, it would have to argued that the distribution of self-citations is not random, for example, that a particular university or group of universities systematically has authors cite themselves while other universities do not. This seems unlikely." Although we agree that self-citation is not a major problem when comparing nations (at overall and field levels), this conclusion is more dubious when it comes to universities. The reason is that universities are rather heterogeneous in their research profiles.

Generally, it is at lower levels of analyses that self-citations represent the most serious problem. In particular this will be the case when assessing individuals and research groups. Here the self-citation rates are likely to show much more variation. Thus, self-citations should preferably be removed before making comparisons. This has also been done in some of the standard indicators by the CWTS-group at the University of Leiden.[10,11] However, other producers of bibliometric studies may not have the necessary data for to carry out such analyses. Also when removing self-citations from individual publications this will undermine the possibility for producing certain types of indicators (e.g., comparisons with average field citation rates).

From another perspective, however, the argument that self-citations do not reflect impact is not that obvious. As an example, consider one paper written by 15 authors. This paper has been cited by another paper written by 10 authors, among whom one (only) was also an author of the cited paper. This count as a self-citation, but the basis for excluding such citations in impact assessments seems less sound. This issue becomes more and more urgent because the average number of authors per publication steadily is increasing. For example, in our national subset the average number of authors increased from 2.6 in 1981 to 4.9 in 1996. In consequence, one should consider if a more narrow definition of self-citation (e.g., counting first-author self-citations only) more adequately captures the citations representing a problem in impact assessments.

Conclusions

In our analysis of Norwegian science we have found that self-citations contribute to a relatively large share of the citations. In the standard indicators based on a three-year citation window on average more than one third of the citations represent author self-citations. However, this share is significantly decreasing when citations are traced for longer periods. In order to reduce the effect of self-citations it would therefore be preferable to use longer citation windows. At aggregated levels (e.g. of nations) we would still argue that self-citation is not a major problem – assuming that self-citations then are being levelled out. Further investigations are, however, required in order to settle this question. At lower levels of analyses self-citations represent a more serious problem. Here, self-citations should preferably be removed before making comparisons. At least, effects caused by self-citations should be carefully considered before using citations as indicators of scientific impact.

*

The author would like to thank Arie *Rip* and Gunnar *Sivertsen* for valuable comments on an earlier draft of this paper.

References

1. T. J. PHELAN, A compendium of issues for citation analysis, *Scientometrics*, 45 (1999) 117–136.
2. S. M. LAWANI, On the heterogeneity and classification of author self-citations, *Journal of the American Society for Information Science*, 33 (1982) 281–284.
3. P. O. SEGLEN, Citations and journal impact factors: questionable indicators of research quality, *Allergy*, 52 (1997) 1050–1056.
4. M. H. MACROBERTS, B. R. MACROBERTS, Problems of citation analysis: A critical review, *Journal of the American Society for Information Science*, 40 (1989) 342–349.
5. L. EGGHE, R. ROUSSEAU, *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Elsevier Science Publishers, Amsterdam, 1990, p. 220.
6. E. GARFIELD, Is citation analysis a legitimate evaluation tool?, *Scientometrics*, 1 (1979) 359–375.
7. R. TAGLIACOZZO, Self-citation in scientific literature, *Journal of Documentation*, 33 (1977) 251–265.
8. S. BONZI, H. W. SNYDER, Patterns of self citation across fields of inquiry, *Proceedings of the ASIS annual meeting*, 27 (1990) 204–207.
9. H. SNYDER, S. BONZI, Patterns of self-citation across disciplines (1980-1989), *Journal of Information Science*, 24 (1998) 432–435.
10. T. N. VAN LEEUWEN, E. J. RINIA, A. F. J. VAN RAAN, *Bibliometric Profiles of Academic Physics Research in the Netherlands.* Centre for Science and Technology Studies. Report CWTS 96-09, Leiden, 1996.

11. H. F. MOED, J. G. M. VAN DER VELDE, *Bibliometric Profiles of Academic Chemistry Research in the Netherlands,* Centre for Science and Technology Studies. Report CWTS-93-08, Leiden, 1993.
12. A. J. NEDERHOF, R. F. MEIJER, H. F. MOED, A. F. J. VAN RAAN, Research performance indicators for university departments: A study of an agricultural university, *Scientometrics*, 27 (1993) 157–178.
13. H. HERBERTZ, Does it pay to cooperate? A bibliometric case study in molecular biology, *Scientometrics*, 33 (1995) 117–122.
14. A. F. J. VAN RAAN, The influence of international collaboration on the impact of research results, *Scientometrics*, 42 (1998) 423–428.
15. E. GARFIELD, *Citation Indexing – Its Theory and Application in Science, Technology and Humanities,* John Wiley & Sons, New York, 1979, p. 245.
16. S. BONZI, H. W. SNYDER, Motivations for citation: A comparison of self citation and citation to others, *Scientometrics*, 21 (1991) 245–254.
17. E. GARFIELD, A. WELLJAMS-DOROF, Citation data: their use as quantitative indicators for science and technology evaluation and policy-making, *Science and Public Policy*, 19 (1992) 321–327.

6. ARTICLE IV. Validation of bibliometric indicators in the field of microbiology. A Norwegian case study.

Aksnes, D. W., T. B. Olsen & P. O. Seglen

Published in Scientometrics 49(1) (2000): 7-22.

Jointly published by Kluwer Academic Publishers, Dordrecht and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 49, No. 1 (2000) 7–22

Validation of bibliometric indicators in the field of microbiology: A Norwegian case study

DAG W. AKSNES, TERJE BRUEN OLSEN, PER O. SEGLEN

Norwegian Institute for Studies in Research and Higher Education (NIFU), Oslo (Norway)

This paper addresses two related issues regarding the validity of bibliometric indicators for the assessment of national performance within a particular scientific field. Firstly, the representativeness of a journal-based subject classification; and secondly, the completeness of the database coverage. Norwegian publishing in microbiology was chosen as a case, using the standard ISI-product *National Science Indicators on Diskette* (NSIOD) as a source database. By applying an "author-gated" retrieval procedure, we found that only 41 percent of all publications in NSIOD-indexed journals, expert-classified as microbiology, were included under the NSIOD-category Microbiology. Thus, the set of defining core journals only is clearly not sufficient to delineate this complex biomedical field. Furthermore, a subclassification of the articles into different subdisciplines of microbiology revealed systematic differences with respect to representation in NSIOD's Microbiology field; fish microbiology and medical microbiology are particularly underrepresented.

In a second step, the individual publication lists from a sample of Norwegian microbiologists were collected and compared with the publications by the same authors, retrieved bibliometrically. The results showed that a large majority (94%) of the international scientific production in Norwegian microbiology was covered by the database NSIOD. Thus, insufficient subfield delineation, and not lack of coverage, appeared to be the main methodological problem in the bibliometric analysis of microbiology.

Introduction

How representative are bibliometric indicators of national research activity within various scientific fields? This is a basic methodological question in bibliometrics. Since such indicators represent a major application of bibliometrics, and may function as a premise for national science policy, it is important that they rest on a valid and representative foundation. In this study two different validity questions are addressed, based on a case study of Norwegian publishing in microbiology: Firstly, how well does a journal-based subject classification represent national performance within a particular scientific field? Secondly, how large is the proportion of the scientific production that is covered by bibliometric databases?

0138–9130/2000/US $ 15.00

A journal-based research field delineation

Defining or delineating a scientific field or subfield is a general problem that does not only arise in the context of bibliometrics. Even among experts there may sometimes be considerable disagreement on how a particular field should be defined, and the increasing interdisciplinarity of scientific research makes it more and more difficult to draw boundaries between the various fields and subfields. In bibliometrics, particular problems arise because many scientific articles – perhaps the majority – deal with complex subject matters that cannot be unequivocally assigned to a single field. For example, one and the same biomedical paper may well be justifiably classified as, e.g., biochemistry, molecular biology, microbiology, immunology or clinical medicine. Because of this, there bound to be overlaps between subfields. Typically, when classifying scientific articles some of the papers will be easy to classify as being within the core of a particular subfield, others will be more difficult, being only weakly related to the subfield or belonging to the borderline area between different subfields.

Despite such general problems, practical working delineations are required in order to do bibliometric research. Ideally, the criteria applied should retrieve all publications belonging to a particular subfield, and at the same time be precise enough to exclude publications irrelevant to the field. Most bibliometric analyses employ journal-based subfield definitions, meaning that all articles in a given journal are assigned to the same subfield.[2, 3] The most widely used database for scientometric purposes, the *Science Citation Index* (SCI) from the Institute for Scientific Information (ISI), classifies its journals into more than one hundred scientific subfields. Other suppliers of scientific data, often based on raw material provided by ISI, have largely retained ISI's subfield designations, but given them a different content by introducing new definitions. For example, Computer Horizons Inc. has developed a subfield delineation based on journal sets derived from journal-to-journal citations.[4] The journal set may, furthermore, even vary between different ISI database products, making comparisons difficult.[5]

Another problem is that the journal set which defines each field may change with time by the inclusion of new journals and the removal of old ones. For some fields, even minor modifications may result in heavy distortions of the apparent national science performance.[6] This problem, and the fact that apparently identical subfield categories may not cover the same set of journals, thus generating incompatibility between bibliometric indicators produced by different institutions, has evoked a discussion of the necessity of standardising bibliometric procedures.[4, 7]

The main problem, however, is related to the very principle of using journals as the units of subfield classification. In the ISI database product *National Science Indicators on Diskette* (NSIOD) the large majority of journals are assigned to a single subfield, although they may contain articles dealing with a relatively broad range of fields. Many articles have multiple subfield affiliations, yet they will be recorded as contributing to just one of these subfields. Thus, some articles will be included in a subfield to which they are only weakly related, whereas other, highly relevant articles will be missing. Some journals are listed under more than one subfield, which does not help much in assigning the proper field to the individual article. Papers in multidisciplinary journals like *Science* and *Nature* are not at all attributed to specific ISI subject categories. In NSIOD, however, these two journals, along with the *Proceedings of the National Academy of Sciences of the USA (PNAS)*, have been reassigned to specific categories by an algorithm which takes into account the category most strongly represented in the paper's title.

Despite the general recognition of these problems, little work has been done to examine how representative the journal-based subfield definitions actually are. Studies by *Lewison*,[8] using a complex assembly of search criteria to retrieve articles from scientifically defined biomedical subfields, showed that only a minor proportion of the articles were included in the specialist journals commonly used to define subfields. A study of the Dutch publication output in physics,[9] comparing the journal category classification of the SCI with the classification of individual publications into subfields in the *Physics Briefs* (PHYS) database, concluded that journal classification yields an incomplete picture of the output of a country.

In the present study, we are introducing a different approach, using scientific authors as a link between articles encompassed by ISI's journal-based subfield definition and articles (by the same authors on similar subjects) classified in the database under different subfield headings. Our investigation has been limited to Norwegian contributions, i.e., articles with a Norwegian address. As a suitable case subfield we have chosen Microbiology, which is a typically complex biomedical discipline that includes basic as well as applied and clinical research. It corresponds to one of ISI's major subject divisions, and should in this context perhaps be regarded as a field rather than as a subfield, yet it is small enough to include less than one hundred Norwegian papers each year, thus manageable from a practical point of view. Microbiology has clear affiliations to other major ISI fields/subfields like Biology & Biochemistry, Pharmacology, Immunology, Molecular Biology & Genetics, and Clinical Medicine, and would, therefore, be expected to typify the problems associated with a journal-based field delineation.

The problem of database coverage

The question of how well bibliometric databases cover scientific production has long been recognised as an important methodological issue in bibliometrics. Historically, SCI was constructed on the basis of Bradford's law, saying that a relatively small number of journals publish the majority of the significant scientific results. Thus, the underlying purpose was to cover the core journals within the different scientific disciplines, but not the peripheral and unimportant journals. An extensive evaluation process is carried out in order to accomplish this purpose, involving an assessment of factors such as coverage, publishing standards, internationality of authorship, English language titles and abstracts as well as journal citation scores.[10]

Nevertheless, it is generally known that several problems exist with respect to the coverage of the ISI database, and in particular on how well it represents the scientific production within a country or scientific subfield. A basic problem is related to the varying traditions and habits regarding publication and citation. Only in fields in which publication in international journals is the major mode of communication can such databases be expected to give a representative picture. For example, a study of the research performance at Australian universities[11] showed that the percentage of the overall scientific production represented in ISI source journals varied considerably among the disciplines. Whereas the chemical, physical, biological, agricultural and medical sciences had more than 75% of their published research output appearing in ISI journals, the corresponding percentages were significantly lower in the earth sciences, mathematics, engineering and information sciences (see also Ref. 12). Furthermore, in the social sciences and the humanities where books are an important publication category, the problem of database coverage is particularly great, although varying between the different subfields.[13–15]

In the present study we have addressed this question in relation to Norwegian microbiology. Based on a mail survey, in which the publication lists of Norwegian microbiologists were collected, the recoveries obtained by various bibliometric approaches have been assessed.

Methods

The commonly used database product *National Science Indicators on Diskette* (NSIOD) provided by ISI was selected as a source of primary data. NSIOD contains aggregated publication and citation counts, classified into 24 different scientific fields/subfields on the basis of journal assignments. In the 1997 edition the field

Microbiology consists of papers (articles, notes, proceedings papers and reviews) from 85 different journals. To test how well this category represents the actual national performance in the field, an "author-gated" retrieval method was applied.

Firstly, we identified the Norwegian subset of articles within the Microbiology category of NSIOD (the 1997 edition, with data covering 1981-1996), by performing a search in the ISI-database *National Citation Report* (NCR). A five-year publication period (1992-1996) was selected. A total of 409 Norwegian papers (i.e., papers with a Norwegian address) were identified by this method (*primary papers*). After correcting for synonymies and spelling mistakes, a total of 1,033 different persons were identified as the authors of these papers (termed *primary authors)*.

Secondly, we used the primary authors as a gate to the microbiological literature outside the NSIOD-defined field, by searching for all 1992-1996 papers in the NCR database with a Norwegian address, authored or co-authored by the 1,033 primary authors. This gave us a sample of *secondary papers*. To eliminate homonymies, we examined the titles, authors' names and institutional addresses of the papers, as well as applying a database of Norwegian researchers (*Forskerpersonalregisteret*). In this limited and transparent material homonymies could be relatively easily identified, and accounted for about 7% of the retrieved articles. With larger material, or in countries with less name diversity, the homonymy problem would be more severe. After the cleansing, we were left with 2,267 articles by 4,049 authors, each paper including at least one primary author.

All articles were then classified on the basis of their titles (and, if necessary, their abstracts) as either microbiology or non-microbiology by a biomedical professional, using a definition of microbiology derived from textbooks and discussions with microbiologists. The articles were also classified into comparable microbiological and non-microbiological subdisciplines. In a mail survey (described below) the definition of microbiology was rated as "good" or better by 72% of the responding scientists. Furthermore, the authors' own classification of their papers as microbiology or non-microbiology was in 92% agreement with the expert classification, indicating that the latter was highly valid.

All authors having publications classified as microbiology were allocated to research groups/clusters on the basis of co-authorship. The most productive author within each group was defined as the group leader. Persons with more than 50% of their articles co-authored with another more productive person were included as group members. Authors sharing articles with several more productive persons, were distributed proportionally on the different groups. On the basis of this method, 149 different microbiology groups were defined, comprising all the authors.

From this identified group structure, a sample of the microbiologists was selected for a mail survey. A questionnaire was sent to three categories of authors: 1) all the persons defined as group leaders, 2) a sample of approximately 50 other productive researchers not defined as group leaders, and, 3) a sample of approximately 50 researchers with only one or two publications within microbiology. This method of selection was assumed to cover a representative sample of Norwegian microbiologists.

In the mail survey the researchers were asked to submit their own complete publication list for the period 1992-96, and to mark the publications they regarded as being within the field of microbiology. In addition, they were asked to answer a few questions on a questionnaire. Among the 240 researchers selected for the survey, 55% responded. This response rate may seem rather low. However, despite the small size of this sample (3% of all authors), the publication lists received covered a relatively large part (64%) of the bibliometrically identified expert-classified microbiology articles. Thus, one may assume a high representativity of the results obtained in the survey.

In the questionnaire the researchers were also asked to add possible missing names of Norwegian microbiologists to an enclosed list of secondary authors with publications classified as microbiology. The respondents identified a total of 53 additional microbiologists. Among these, 22 were represented in NSIOD, with 37 microbiological and 75 non-microbiological publications. Thus, the number of microbiology articles was increased by 4%. This indicated that our author-gated retrieval method had been relatively successful in identifying the large majority of the international articles published by Norwegian microbiologists.

Finally, we analysed the material from the mail survey, sorted the publications by type (articles in international journals, national journals, book chapters, etc.), classified them into scientific subfields, and compared the authors' publication lists with the material recovered bibliometrically.

Results

Poor coverage, but high relevance of the journal-based field delineation

Among the 2,267 articles (by 4,049 authors) obtained by our author-gated retrieval procedure, 976 (by 1,943 authors) were expert-classified as microbiology, and 1,291(by 2,513 authors) as non-microbiology. This approach thus uncovered 580 microbiology papers in addition to the 409 papers included in the original NSIOD Microbiology field. Among the latter, only 13 papers were reclassified by the expert as being non-microbiological, i.e., the relevance of the journal-based field definition was extremely high (97%). The journal set that defines the field would thus seem to represent a hard core of microbiological specialist journals. In contrast, since a mere 41% of the microbiological articles in the database were included in the defined Microbiology field, the degree of coverage could only be described as moderate. Apparently the majority of microbiological research papers are published outside the core journals.

The 580 additional microbiology papers identified through the author-gated retrieval method were published in 212 different journals, indexed in 16 of the standard NSIOD fields/subfields. Prominent among these other fields were Clinical Medicine (251 articles), Plant & Animal Science (141), Immunology (127), Agricultural Sciences (45), Biology & Biochemistry (28) and Molecular Biology & Genetics (27).

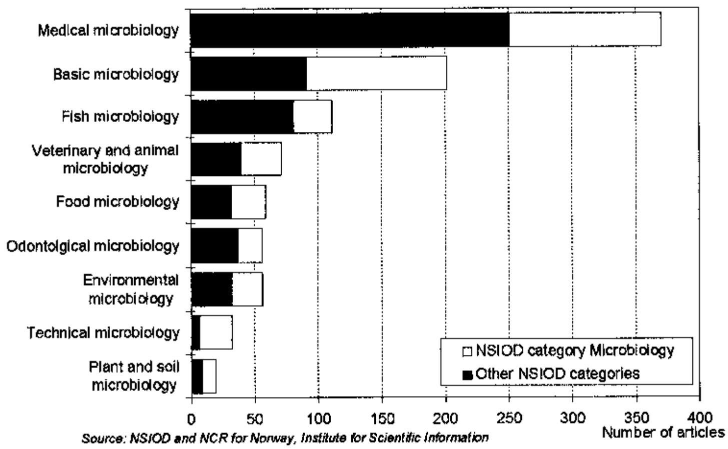


Fig. 1. Norwegian articles in microbiology, 1992-1996, by subfield. Author-gated retrieval

Differences in coverage between microbiological subdisciplines

The microbiology papers were subclassified into nine microbiological subdisciplines as shown in Fig. 1. In the author-gated retrieval a large fraction of the papers (38%) were in the relatively broad category Medical microbiology, which included clinically oriented subjects like infectious diseases, infection-related immunology, and microbial epidemiology. Basic microbiology, i.e., studies of the basic properties of micro-organisms (biochemistry, cell biology, molecular biology and genetics), accounted for

21% of the articles. These findings for microbiology are in accordance with the previously established research activity profile for Norway, which shows a much higher activity in clinical medicine than in basic biomedicine.[6] The low overall percentage of basic microbiological articles is probably in part a reflection of the rather low Norwegian investments in basic science, compared, e.g., to Sweden.[16] Fish-related microbiology accounts for as many as 11% of the articles, presumably another characteristic feature of Norway's research profile.

Figure 1 reveals considerable variation between the microbiological subdisciplines with regard to coverage within NSIOD's defined microbiology field. Technical microbiology is particularly well covered (78%), possibly because other NSIOD fields are relatively lacking in journals suitable for this subject, but the number of papers in this subdiscipline (and in some of the others) is really too small for any firm conclusions. Basic microbiology has a reasonably good coverage of 55%, reflecting the fact that probably all important basic microbiology journals are included in the field-defining journal set. In contrast, medical and fish microbiology are poorly covered (32% and 27%, respectively). Apparently, microbiologists in these fields tend to publish in more general or applied journals which are not defined as microbiology journals in NSIOD. The fact that the journal coverage varies between subdisciplines may have important consequences for the recording of national science performance: Norway, with its low emphasis on basic science, would be expected to have a relatively low fraction of its microbiological research included in the NSIOD-defined Microbiology field.

Citedness of microbiology articles

The NCR database includes citation data, which has allowed us to count the citations made to the NSIOD papers retrieved in our study. By the end of 1996, the 976 microbiology articles had received a cumulated total of 4,984 citations, giving an average of 5.1 citations per article. The articles inside the NSIOD-defined microbiology field were more highly cited (5.7, i.e., identical to the world average) than the microbiology articles in other NSIOD fields/subfields (4.6, i.e., 20 per cent below the world average). This difference was statistically significant at the 5% confidence level, according to Student's t-test. The 1,291 non-microbiology articles had been cited 7,390 times, i.e., 5.7 citations/article.

An analysis of subdisciplines revealed appreciable differences in citedness, but the number of articles in many of these disciplines was obviously too small to provide reliable citation data. We have, therefore, looked at the medical and basic subdisciplines separately, but combined the other subdisciplines into a single "other" category. As shown in Fig. 2, basic microbiology had the highest overall citedness, medical microbiology was intermediate, and the other microbiological disciplines had the lowest citedness. The citation rates inside and outside the NSIOD Microbiology field were quite similar for basic microbiology, whereas medical microbiology and the other subdisciplines were somewhat more highly cited inside the defined Microbiology field. We also collected the citation data for the non-microbiology papers. Strikingly, these articles were most highly cited in basic biomedicine (even more so than the corresponding microbiology articles), moderately in general/clinical medicine, and considerably less in the other scientific disciplines. These differences probably reflect the well-established fact that the basic sciences tend to have higher citation rates than the more applied sciences, due to a largely unidirectional flow of references between these two major divisions of science.[17-19] Within the basic biomedical sector, microbiology would seem to be a relatively application-oriented subject compared to the rest of the sector, and hence receive fewer citations. In contrast, among the practically oriented disciplines included in the "other" category, microbiology would stand out as a relatively basic subject, and hence be more cited than the rest of the category.

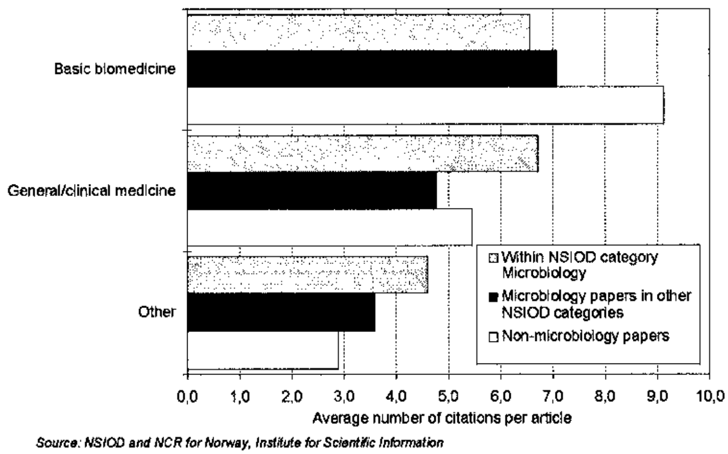


Fig. 2. Average number of citations of Norwegian articles in microbiology, 1992-1996, by field. Author-gated retrieval

The citedness of the Microbiology field as a whole will, therefore, depend upon the relative contribution from basic microbiology compared to the more applied microbiological disciplines. Since NSIOD's journal-based field definition is somewhat biased in favour of basic microbiology, a more inclusive field delineation does in fact reduce the recorded overall citedness of the field from 5.7 to 5.1.

Analysis of authors' publication lists, obtained by mail survey

On the basis of a mail survey, 120 publication lists (from 93 primary and 27 secondary authors) were received. The lists included as many as 64% (625/976) of the microbiology articles in the secondary material, indicating that a very representative sample had been obtained. An additional 150 articles in the reference lists were expert-classified as microbiological, among which 105 were recorded in NSIOD. One-fifth of the latter was authored by secondary authors only, and would thus have been missed by the primary author-gated retrieval. Why the other had been missed was not investigated in detail, but several factors (lack of a Norwegian address; a document type not recorded in NSIOD, etc.) would appear to be involved. With as few as 45 non-NSIOD articles, the overall database coverage was an impressive 94% for international journal articles in microbiology, with high values for all subfields. For bibliometric analysis of international scientific article production, a bibliometric database like NSIOD would thus give a very representative picture of the microbiology field.

The results from the mail survey are presented in Table 1. Only publications classified by the expert as microbiology have been included in the analysis. It should be emphasised, however, that most of the respondents had also published in other fields, in particular medicine and basic biomedicine; only 59% of their overall scientific production was in fact within microbiology.

In addition to international journal articles/reviews, the authors' reference lists included 107 microbiological non-journal publications (books, book chapters, proceeding papers) as well as 149 microbiological publications (of all the above-mentioned categories) in Norwegian. Abstracts, posters, book reviews, editorials, reports, etc., were not counted. Although there is no standard solution to the question of which publication types ought to be included in bibliometric analyses, the selection made here would not seem unreasonable. By inclusion of the counted non-journal and Norwegian publications in the overall microbiology production, the database coverage fell to 71%, with values as low as 40-50% for the subfields Veterinary & animal microbiology, Plant, soil & technical microbiology and Food microbiology, and 65% for Fish microbiology. For these subfields, a bibliometric analysis would thus strongly underrepresent the overall volume of the scientific production. Research within these subfields often has an applied character, and an emphasis on imparting knowledge to a non-scientific public, for example, through trade journals. Publication in national journals is particular prominent within Veterinary and Medical microbiology, with many articles in *The Journal of the Norwegian Veterinary Association* (*Norsk veterinærtidsskrift*) and *The Journal of the Norwegian Medical Association (Tidsskrift for Den norske lægeforening)*, respectively.

As in the general bibliometric analysis, the microbiological international journal articles retrieved from the publication lists were poorly covered (31%) by the NSIOD-defined Microbiology field. Taking all counted publication types into account, the coverage was only 23%, with values for some subfields (e.g., Fish microbiology) as low as 10%. A journal-based field definition is thus not very useful for the assessment of national performance within specific scientific subfields.

Table 1
Microbiology publications from authors' publication lists (mail survey): relative representation in the NSIOD database

	International journal articles				Proceedings, books, etc.		Publications in Norwegian		All scientific publications		
Microbiology subfield	No.	% of total	% NSIOD coverage	% NSIOD Microb.	No.	% of total	No.	% of total	No.	% NSIOD coverage	% NSIOD Microb.
Microbiology, all fields	775	75	94	31	107	10	149	14	1031	71	23
Basic	164	85	99	49	27	14	3	2	194	84	42
Medical	329	80	95	27	32	8	49	12	410	76	22
Odontological	45	94	98	31	3	6	0	0	48	92	29
Food	47	53	96	43	11	13	30	34	88	51	23
Veterinary & animal	50	48	84	22	3	3	52	50	105	40	11
Environmental	35	90	97	23	4	10	0	0	39	87	21
Fish	92	77	84	13	12	10	15	13	119	65	10
Plant, soil & technical	13	46	100	15	15	54	0	0	28	46	7

Our author-gated bibliometric retrieval yielded 976 microbiological journal articles, 625 of these (64%) were also recovered in the reference list sample along with 150 additional articles. By extrapolation, a 100% list recovery should have given 234 additional articles, i.e., a total of 1,210 journal articles. By adding the 37 articles published by microbiologists identified through the mail survey, the resulting estimated grand total of 1,247 microbiology articles would presumably reflect rather closely the overall Norwegian production of journal articles within the microbiology field during the study period. The 78% retrieval obtained by author-gating would seem relatively acceptable, at least compared to the 32% recovery (396 microbiology articles) obtained using NSIOD's journal-based field definition.

Discussion

Our study has shown that a large majority of the international scientific publications in Norwegian microbiology is covered by the NSIOD database. At a general level, NSIOD thus includes the journals in which the majority of the microbiology papers is published. Only a few journals are missing and these are probably rather insignificant as far as microbiology is concerned. However, the microbiology production is seriously underestimated in the NSIOD category Microbiology, most of the production being assigned to other fields.

Bibliometric databases, and in particular the one maintained by ISI, are often regarded as biased in favour of English-speaking countries (see, e.g., Ref. 20). This factor does not seem to be of significance when it comes to Norwegian microbiology. In part this may reflect that microbiology is a discipline in which important scientific communication is channelled through international, mostly English-language journals – as typical for most medical and scientific disciplines. In fact, a former study of scientific publishing at Norwegian universities showed that the publication output in the biomedical sciences consisted of more than 80 percent English-language literature – a higher share than in any other academic field.[20] Thus, this may suggest that the biomedical sciences would be among the most suited disciplines for bibliometric analyses.

Despite the crucial significance of international papers, an important question would be how national publications and other kinds of "grey literature" should be assessed. As a general point, it might be argued that all contributions to the research front in microbiology – at least of any significance – would be expected to be communicated through international journals. This means that a bibliometric database covering this literature would be well suited for assessing contributions to scientific development. However, producing forefront knowledge is not the only purpose of science. For instance, in some fields, communication with a non- or semi-scientific public may be an important additional responsibility. In fish microbiology, imparting new knowledge to fish farmers represents such an example. The role of trade journals in fields like agricultural sciences has been analysed in former studies.[22] Although this kind of literature sometimes may be of dubious scientific quality, and is not directed towards the international scientific community, it is important to be aware of the function it represents in the research system. Therefore, in fields where other publication types than international journal articles are more common, the limitations of standard bibliometric indicators should always be recognised, particularly if such indicators are used in science policy.

The main problem shown in our validation study is, however, related to bibliometric subject classification. At a general level serious questions can be asked about the principles that underlie ISI's assignment of journals. This is particularly evident from the fact that in the standard version of NSIOD, the subfield Microbiology consists of 85 different journals, while 72 journals are included in the subfield Microbiology in the SCI Expanded. However, only 45 journals are identical in the two products. That is, 27 of the journals in SCI Expanded are not included in NSIOD, while 40 of the journals in NSIOD are not included in SCI Expanded.

In any case, the fact that less than one-half of the Norwegian publications within microbiology were covered by NSIOD's journal-defined Microbiology field, suggests that journals do not provide a very suitable basis for field delineation, at least not for in-depth studies of specific fields. The shortcomings of journal-based field delineations have also been pointed out in Refs 5, 8, 20, and 23. *Lewison*[8] showed that biomedical subfields defined by a set of specialist journals covered only a minority of the articles in the actual subfield, most articles being published in general journals. A standard bibliometric analysis based on the NSIOD classification would thus seriously underestimate the Norwegian productivity in microbiology, while overestimating it in adjacent, more broadly defined fields. This would be the case for other countries as well, suggesting that comparisons between countries might still be valid. However, since the journal set that defines the field seems to be dominated by basic core journals, the article retrieval will be biased in favour of basic microbiology. The measured national productivity in the field would, therefore, be significantly influenced by the ratio of basic vs. applied microbiology, a ratio that would be expected to differ between nations. Norway, with its low investments in basic research,[6, 16, 24] would be expected to present a particularly poor picture compared to other countries. Not surprisingly, Norway ranks far behind comparable countries like Sweden and Denmark on the bibliometric productivity statistics for microbiology (which are based on the NSIOD database[24]).

With citations, the situation is different. Since basic science is more highly cited than applied science, a restrictive field definition that tends to exclude applied microbiology will improve the recorded citation impact of a nation with an application-directed emphasis. This is the case for Norway: the "official" Norwegian citation impact is in fact comparable to that of Sweden and Denmark.[24] Obviously, it is important to have full information about the total activity, the total citedness and the subdiscipline structure within a field or subfield in order to draw the appropriate conclusions regarding national scientific performance.

It should be emphasised that the field delineation problem is basically a distributional one: with a fixed data set such as NSIOD, one field's loss will be another field's gain. The overall productivity or citedness of a nation, across all fields, should thus not be affected by an improper field assignment. On the other hand, in a field-to-field comparison, even the net result could be affected; for example, the transfer of one field's most highly cited articles to another, generally more highly cited, field could actually reduce the mean citedness of both fields. An additional complication is the assignment of multidisciplinary papers or journals to just one field. Without the knowledge of the degree of bias in such assignments, it can be difficult to tell whether a field is underrepresented or overrepresented. The solution is a fractional field assignment, but this would be technically cumbersome.

It has been argued that a journal-based field delineation may be more appropriate for broad fields than for narrow subfields,[25] because the broad fields could include both the specialist journals and the more general journals which actually contain most of the specialised articles.[8] This may be the case in some of the natural sciences, but within biomedical research the fields overlap (in terms of journals) to such an extent that the problem of representativeness remains, and should be kept in mind, even with a broad field definition. Insulated subfields, completely dominated by a few specialist journals would, of course, be well suited for a journal-based field delimitation,[26] but such subfields would be uncommon – microbiology is certainly no example.

Obviously, no mechanical system will be able to classify correctly all papers. As indicated by our study, a search algorithm based on researcher names was highly effective, with an estimated 78% retrieval of the relevant articles. However, the specificity was moderate, since a relatively large part (57%) of the researchers' output was outside the study field. Alternative methods of field delineation that have been tried include the use of department names or cognitive terms from institutional addresses, but this particular strategy was not found to be very useful.[25] A more promising approach would be to access the article contents directly, through a search for keywords, title words, or various combinations of these.[5, 8] Search algorithms in which a set of core journals is combined with keywords and title words have proved to be particularly effective.[8] Some literature databases, like *Chemical Abstracts* and *Medline*, do in fact supply each article with indexed keywords. Our author-gated retrieval procedure would seem well suited for a combinatorial approach in which the present expert classification is replaced by a keyword-based search algorithm to produce a fully computerised bibliometric method for analysis of scientific fields.

*

Some preliminary results from this project were presented at the 7th International Conference on Scientometrics and Informetrics, Colima, Mexico, 5-8 July 1999.[1]

References

1. D. W. AKSNES, T. B. OLSEN, P. O. SEGLEN, Inadequacy of a journal-based research field delineation. Incomplete recovery of Norwegian microbiology articles in ISI's Microbiology field. In: C. A. MACÍAS-CHAPULA (Ed.) *Proceedings of the Seventh Conference of the International Society for Scientometrics and Informetrics*, Colima, Mexico, July 5-8, 1999.
2. J. F. MIQUEL, T. OJASOO, Y. OKUBO, A. PAUL, J. C. DORÉ, World science in 18 disciplinary areas: Comparative evaluation of the publication patterns of 48 countries over the period 1981-1992, *Scientometrics*, 33 (1995) 149–167.
3. T. BRAUN, W. GLÄNZEL, H. GRUPP, The scientometric weight of 50 nations in 27 science areas, 1989-1993. Part 1. All fields combined, mathematics, engineering, chemistry and physics, *Scientometrics*, 33 (1995) 263–293.
4. H. F. MOED, Differences in the construction of SCI based bibliometric indicators among various producers: A first overview, *Scientometrics*, 35 (1996) 177–191.
5. I. GÓMEZ, M. BORDONS, M. T. FERNÁNDEZ, A. MÉNDEZ, Coping with the problem of subject classification diversity, *Scientometrics*, 35 (1996) 223–235.
6. G. SIVERTSEN, *Norsk forskning på den internasjonale arena: En sammenligning av 18 OECD-lands artikler og siteringer i Science Citation Index 1973-86.* NAVFs utredningsinstitutt, Oslo, Rapport 1/91.
7. W. GLÄNZEL, The need for standards in bibliometric research and technology, *Scientometrics*, 35 (1996) 167–176.
8. G. LEWISON, The definition of biomedical research subfields with title keywords and application to the analysis of research output, *Research Evaluation*, 6 (1996) 25–36.
9. E. J. RINIA, C. DE LANGE, H. F. MOED, Measuring national output in physics: Delimitation problems, *Scientometrics*, 28 (1993) 89–109.
10. E. GARFIELD, The ISI database: The journal selection process. Internet: URL: http://www.isinet.com/whatshot/essays, 1997.
11. P. BOURKE, L. BUTLER, Publication types, citation rates and evaluation, *Scientometrics*, 37 (1996) 437–494.
12. H. F. MOED, W. J. M. BURGER, J. G. FRANKFURT, A. F. J. VAN RAAN, The use of bibliometric data for the measurement of university research performance, *Research Policy*, 14 (1985) 131–149.
13. A. J. NEDERHOF, R. A. ZWAAN, R. E. DE BRUIN, P. J. DEKKER, Assessing the usefulness of bibliometric indicators for the humanities and the social and behavioural sciences: A comparative study, *Scientometrics*, 15 (1989) 423–435.

14. U. SCHOEPFLIN, Problems of Representativity in the Social Sciences Citation Index. In: P. WEINGART, R. SEHRINGER, M. WINTERHAGER (Eds) *Representations of Science and Technology. Proceedings of the International Conference on Science and Technology Indicators,* Bielefeld, Germany, 10-12 June 1990. DSWO Press, Leiden University, 1992.
15. D. HICKS, The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, *Scientometrics,* 44 (1999) 193–215.
16. Norges Forskningsråd, *Det Norske Forskningssystemet – Statistikk og Indikatorer.* Oslo, 1997.
17. F. NARIN, G. PINSKI, H. H. GEE, Structure of the biomedical literature, *Journal of the American Society for Information Science,* 27 (1976) 25–45.
18. G. FOLLY, B. HAJTMAN, J. I. NAGY, I. RUFF, Some methodological problems in ranking scientists by citation analysis, *Scientometrics,* 3 (1981) 135–147.
19. P. O. SEGLEN, Citations and journal impact factors: questionable indicators of research quality, *Allergy,* 52 (1997) 1050–1056.
20. A. J. NEDERHOF, Delimitation of a medical research topic: interaction with experts in selecting a database and constructing a search strategy, *Research Evaluation,* 1 (1991) 149–154.
21. S. KYVIK, *Productivity in Academia. Scientific Publishing at Norwegian Universities*, Norwegian University Press, Oslo, 1991, p. 81.
22. A. J. NEDERHOF, R. F. MEIJER, Development of bibliometric indicators for utility of research to users in society: measurement of external knowledge transfer via publications in trade journals, *Scientometrics,* 32 (1995) 37–48.
23. M. BORDONS, S. BARRIGÓN, Bibliometric analysis of publications of Spanish pharmacologists in the SCI (1984-89). II. Contribution to subfields other than "Pharmacology & Pharmacy" (ISI), *Scientometrics,* 25 (1992) 425–446.
24. T. B. OLSEN, *Norsk forskning i internasjonale tidsskrifter: Sammenligning med andre land belyst ved bibliometriske makrodata.* NIFU, Oslo, Rapport 1/98.
25. R. E. D. DE BRUIN, H. F. MOED, Delimitation of scientific subfields using cognitive words from corporate addresses in scientific publications, *Scientometrics,* 26 (1993) 65–80.
26. A. SCHUBERT, T. BRAUN, Cross-field normalization of scientometric indicators, *Scientometrics,* 36 (1996) 311–324.

Received March 20, 2000.

Address for correspondence:
DAG W. AKSNES
Norwegian Institute for Studies in Research and Higher Education (NIFU)
Hegdehaugsveien 31, N-0352 Oslo (Norway)
E-mail: dag.w.aksnes@nifu.no

7. ARTICLE V. Peer reviews and bibliometric indicators. A comparative study at a Norwegian university.

Aksnes, D. W. & R. E. Taxt.

Published in <u>Research Evaluation</u> 13(3) (2004): 33-41.

Research Evaluation, volume 13, number 1, April 2004, pages 33–41, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Evaluation correlation

Peer reviews and bibliometric indicators: a comparative study at a Norwegian university

Dag W Aksnes and Randi Elisabeth Taxt

This study investigates the relationship between bibliometric indicators and the outcomes of peer reviews. Based on a case study of research groups at the University of Bergen, Norway, we examine how various bibliometric indicators correlate with evaluation ratings given by expert committees. The analysis shows positive but relatively weak correlations for all the selected indicators. Particular attention is devoted to the reasons for the discrepancies. We find that shortcomings of the peers' assessments, of the bibliometric indicators, as well as lack of comparability, can explain why the correlation was not stronger.

Dag W Aksnes is at the Norwegian Institute for Studies in Research and Higher Education (NIFU), Hegdehaugsveien 31, NO-0352 Oslo, Norway; email: Dag.W.Aksnes@nifu.no Randi Elisabeth Taxt is at the Faculty of Mathematics and Natural Sciences, University of Bergen, Pb. 7800, NO-5020 Bergen, Norway.

CITATIONS HAVE INCREASINGLY been applied as indicators in research assessments. Nevertheless, the question concerning the validity of citations as performance measures is debated and controversial. During recent decades a large number of studies have analysed how citation indicators correlate with peer reviews. In these studies judgements by peers have been typically regarded as a kind of standard by which citation indicators can be validated. The basic assumption is that one should find a correlation if citations legitimately can be used as indicators of scientific performance.

Generally, most of the studies seem to have found an overall positive correspondence although the correlations identified have been far from perfect and have varied among the studies. The studies also differ in methodology and levels of investigation. Luukkonen (1991), for example, found a tendency for citation counts to correlate roughly with peer ratings. In a more recent study Rinia *et al* (1998) found that various citation indicators correlated significantly with peer ratings of research programmes in condensed matter physics. Similarly, Oppenheim (1997) identified strong positive correlations between citation indicators and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology — although his conclusions were challenged by Warner ((2000). Other studies have found a correlation between citation counts and other measures of research impact or scientific recognition. For example, Cole and Cole (1973) found such a correlation in respect to Nobel prizes, honorific awards and reputational ability (see also works by Garfield, e.g. Garfield and Welljams-Dorof, 1992).

 0958-2029/04/010033-09 US$08.00

This paper represents an additional contribution to this issue. The Research Council of Norway has recently initiated several large-scale evaluations of Norwegian research within the natural sciences (Research Council of Norway 1997, 1998, 2000a, 2000b, 2002a, 2002b). These evaluations were carried out by international expert panels but did not involve professional bibliometric analyses (only at an overall national level). It was therefore of particular interest to see how the outcome of the evaluations corresponded with bibliometric indicators and performance measures. In order to answer this question we made a case study of the University of Bergen, which is one of the four Norwegian universities that were included in the evaluations (in addition to some public research institutes). We analysed how peer ratings of the scientific performance of research groups correlated with various bibliometric indicators. Citation indicators were given the main focus in the study, but we also analysed other kinds of bibliometric indicators. The comparative study presented here was carried out as part of a larger project aiming at monitoring the research activities at the Faculty of Mathematics and Natural Sciences at the University of Bergen.

On the basis of findings in former studies, we would have expected to find a significant correlation between peer assessments and bibliometric indicators. However, such a comparative study relies on two assumptions that are not self-evident. First, peer reviews and bibliometric indicators can be expected to correlate only if the two focus on the same aspect of the scientific performance. Second, judgements made by peers may not necessarily be considered as the 'truth' to which bibliometric indicators should correspond – the peers may be biased or mistaken in their assessments. In our study both issues have been taken into consideration in order to explain the results. Particular attention was given to those cases where judgements on the basis of bibliometric results did not correspond with the views of the peers: What might explain these divergent results?

We think such a study is important in order to understand what the indicators measure, their limitations and usefulness as evaluation tools. In this way the study gives further knowledge of the potential and limitations of bibliometric indicators in research assessments.

Peer evaluation – some basic information

Peer evaluations were carried out at national level and in different years: chemistry in 1997, the earth sciences in 1997/1998, biology/biomedicine in 1999/2000, physics in 2000, mathematics and information and communication technology in 2001/2002. Each discipline was evaluated by a panel consisting of seven to nine international experts (biology/biomedicine by three different panels). The panels usually visited Norway for one week, either involving site visits or meetings in a hotel at Oslo airport. Prior to these meetings the panels received various background material and information (for example, data on each department's human resources, financial support and number of students, self-evaluations from each department and CVs of the evaluated scientists, including their lists of publications during the previous five years).

The panels were requested to assess the strengths and weaknesses of each department and research groups with respect to scientific activity and quality, training and mobility, international and national collaboration and organisation (strategy, resources, equipment, personnel and environment). Most of the evaluations adopted a grading system in which each group was given an overall mark: outstanding (excellent) — very good — good — fair — poor (unacceptable). For example, 'outstanding' was typically defined in the following way: 'Research at a very high international level; of great interest with broad impact and with publications in international leading journals; the researchers are among the leaders of the field.'

The field evaluations have been given much attention by the research communities at national and institutional levels. For example, the results have been presented in the national press and there have been a considerable number of readers' letters in university newspapers. The departments evaluated were also given the opportunity to respond to the evaluation reports. The reception has been rather diverse. Some, particular the policy makers, have regarded the main conclusions of the evaluations as very valuable and useful. Some have been more reticent or critical, arguing that the evaluations give a misleading picture of the research performance, are based on inadequate methodologies and contain factual errors. Others have argued that the evaluations 'only told us what everybody already knew' and therefore were not worth the effort and money. The evaluations have also been criticised for being far too superficial, particularly that the panels have not spent enough time to provide credible judgements of the research performance at group levels.

Methods

Analysing the peer evaluations

As one part of the study we examined the peer evaluations of the departments and research groups at the Faculty of Mathematics and Natural Sciences at the University of Bergen. We found that the evaluations differed considerably in structure and focus. Usually, the panels examined the individual research groups in each department, but in a few cases they considered the institutes only at an overall level. Furthermore, most of the evaluations focused on the productivity, scientific quality and relevance of the research of the departments/groups — but with varying emphasis. In one case the panel appreciated its inadequacy with respect to fairly evaluating the scientific significance of the research project being conducted. Therefore this panel focused mainly on organisation, including effective use of resources, collaboration and so forth. The panels also differed in the way they used the rating system and in the importance they attached to the different factors in the awarding of the overall marks.

Because of this diversity we decided to limit our analysis. Considering our purpose of comparing citations and other bibliometric measures with peer ratings, we selected only those groups that were evaluated by the panels in respect of their research profile and results (i.e. scientific quality, significance or impact), and in which this performance was rated, directly or indirectly (through a mark or through explicit statements such as 'the scientific quality of the research is very good'). For example, we excluded those groups that were assessed only in terms of organisation. We also excluded the entire evaluation of physics because the ratings of this evaluation were based mainly on simple productivity measures. Furthermore, we decided to exclude the groups having fewer than six ISI-indexed articles during the time period (because of the poor statistical significance that can be attached to citation indicators in such cases). Finally, we excluded a few groups because they received divergent assessments. In total we were left with 34 research groups.

The bibliometric analysis

As a basis for the project we used the database Institutional Citation Report (ICR), University of Bergen, provided by Thomson-ISI. In order to identify a set of publications as completely as possible, we supplied Thomson-ISI with a list of the more uncommon addresses at the university, which would otherwise be difficult to identify. The search for publications was limited to the six-year period 1995 to 2000 but with citations including 2001. We then identified the names of the tenured scientific personnel working at the Faculty of Mathematics and Natural Sciences in 2000. Here we used the lists of employees, and we consulted each department and asked for information on the group structure applied in the evaluations. The evaluation focused mainly on tenured scientific personnel, and in the bibliometric analysis we included only those articles authored or co-authored by these persons. For each person we made a search for publications in the ICR-database (counting regular articles, notes, letters, proceedings papers and reviews). Through a mail survey each tenured scientist was asked to verify the papers identified in the ICR-database and complete the lists with missing publications from the period chosen. Research fellows and personnel working on external grants were also identified, but only at the departmental level, and they were not asked to verify the papers identified in ICR.

The responses from the tenured scientific personnel were analysed in order to assess the gaps between the total production at the faculty and the publications retrieved from the ICR-database. We then used the ICR-database to calculate different bibliometric indicators for each group. As a basis for the comparative analysis we selected the following five indicators, which represent some of the main indicators that are frequently used in bibliometric performance analyses (see e.g. Moed and van der Velde, 1993; Vinkler, 1998; VSNU, 1996):

1. Number of papers per tenured scientific personnel (fractionalised for co-authorship)
2. Number of citations per person
3. Relative citation rate
4. Relative subfield citedness
5. Relative publication strategy

For the citation indicators we collected the total number of citations for each paper from year of publication through 2001. As a basis for the indicators we used different reference values.

For Indicator 3, relative citation rate (cf. Schubert and Braun, 1986), we used the mean citation rate of the group's journal package, calculated as the average citation rate of the journals in which the group has published, taking into account both type of paper and year of publication (using the citation window from year of publication through 2001). For example, for a letter published in a particular journal in 1998 we identified the average citation rates (1998–2001) of all the letters published by this journal in 1998. ISI refers to this average as the *expected citation rate* (XCR), and it is included as a bibliometric reference value for all publications indexed in ICR. For each group we then calculated the mean citation rate of its journal package, with the weights being determined by the number of papers published in each journal/year. Indicator 3 was then calculated as the ratio of the average citation rate of the group's articles to the average citation rate of its journal package. For example, an index value of 1.1 would mean that the group's articles are cited 10% more frequently than 'expected' for articles published in the particular journal package.

A similar method of calculation was adopted for Indicator 4, relative subfield citedness (cf. Vinkler, 1986, 1997). Here, as a reference value, we used the mean citation rate of the subfields in which the research group has published. This reference value was calculated using the bibliometric data from the Journal Performance Indicator (JPI) database. This is a bibliometric database provided by Thomson-ISI in which the journals are assigned to more than 250 subfield categories. Using this database it is possible to construct a rather fine-tuned set of subfield citation indicators. The groups are usually active in more than one subfield (i.e. the journals they publish in are assigned to different subfields). For each group we therefore calculated weighted averages with the weights being determined by the total number of papers published in each subfield/year.

One problem with this procedure is caused by the multiple assignments of journals. That is, in ISI's classification system some journals are assigned to more than one subfield. In order to handle this problem we used the average citation rates of the respective subfields as a basis for the calculations for the multiple assigned journals.[1] Indicator 4 was then calculated as the ratio of the average citation rate of the group's articles to the average subfield citation rate. In this way, the indicator shows whether the group's articles are cited below or above the world average of the subfield(s) in which the group is active.

We also calculated the relative publication strategy (cf. Vinkler, 1986, 1997) of the groups (Indicator 5). This indicator was calculated by dividing the average citation rate of the journals in which the group's articles were published by the average citation rates of the subfields covered by these journals. Thus, if this indicator exceeds 1.00 one can conclude that the group publishes in journals with a relatively high impact.

We then compared the bibliometric indicators for each group with the ratings from the peer evaluations. In order to undertake correlation analyses the marks were converted to corresponding numbers. The mark 'poor' was converted to the numerical value of 0, 'Fair was converted to the numerical value of 1 and so forth. For the comparative analysis we used the Pearson's correlation coefficient.[2]

It should be noted, however, that the time period applied in the bibliometric study differs somewhat from those used in the evaluations. The panels usually considered the research activities during the preceding five-year period. This means that the period analysed in some of the evaluations (primarily for chemistry and earth sciences) only partly overlapped with that analysed in the bibliometric survey. Still, we do not think this represented a major problem since the panels also considered current research and future research plans in the evaluations – research that would be published at a later stage. We did, however, check as to how our bibliometric indicators would be affected by changing the time window.

It is also worth noting that in the citation indicators the oldest publications will have relatively more weight than the recent publications. This is due to the fact that the 1995 publications, for example, will have assembled citations over a longer time period than articles published in 2000. Nevertheless, our method has some advantages compared to the alternatives. In particular, it reduces the problem of the poor reliability of citations as indicators when very short time periods are considered. The method adopted here is also quite commonly applied in bibliometric performance analyses (see e.g. Moed and van der Velde, 1993; van Raan, 1996).

Our method has some advantages; it reduces the problem of the poor reliability of citations as indicators when very short time periods are considered

Results

In total we identified approximately 2,900 ICR-indexed papers published by researchers at the faculty between 1995 and 2000. Corrected for internal co-authorship this corresponded to 2,034 unique articles. Most (74%) of the tenured scientific personnel responded to the mail survey in which they were asked to list publications not identified in the ICR-search. It appeared that we had retrieved the large majority of the articles published in international journals. There were nevertheless variations among the different departments. The poorest coverage was at the Department of Botany in which the ratio between the number of ICR-identified articles and the missing publications was 1:0.4. The Department of Mathematics and the Department of Informatics also reported relatively high numbers of missing publications (ratio: 1:0.3). On the other hand the list of ICR-publication was almost complete for the Department of Molecular Biology (ratio 1:0.0) and Department of Microbiology (1:0.1). Through the mail survey we also received information on publications in other media than international journals. These results will not be presented here. However, it is worth noting that publication in proceedings was particularly common at the Department of Informatics.

The varying ICR-coverage should be taken into account when interpreting the results of the comparative analysis. Generally, the coverage of the ICR-database must be regarded as fairly good, and the majority of the missing articles appeared in peripheral and obscure journals.

In the comparative analysis we found that the bibliometric indicators and the panel ratings generally showed quite different pictures of the scientific performance at group levels. Using data for the 34 research groups we undertook regression analyses between the panel ratings and the different bibliometric indicators. The correlation coefficients are shown in Table 1.

As may be seen, we found weak positive correlations for all indicators. A correlation coefficient of 0.34 was found for the productivity indicator (number of papers per person fractionalised for co-authorship). For the three citation-based indicators the correlation coefficient varied between 0.24 and 0.46. The weakest correlation was found when the panel ratings were compared with the relative

Table 1. Pearson's correlation coefficients for different bibliometric indicators and panel ratings, using data for 34 research groups

Bibliometric indicator	Correlation coefficient
1. Number of papers per person (fractionalised for co-authorship)	0.34
2. Number of citations per person	0.31
3. Relative citation rate	0.24
4. Relative subfield citedness	0.46
5. Relative publication strategy	0.48

citation rate indicator (i.e. citedness relative to the group's journal package). A slightly higher coefficient was found for the indicator measuring the average number of citations per person. The highest coefficient among the three was found for the indicator in which the citations are measured relative to fields (relative subfield citedness). Finally, we compared the peer ratings with the journal profile of the groups (relative publication strategy). Here we found a correlation coefficient of 0.48. With the exception of the relative citation rate indicator, the correlations were statistically significant at the 0.05 level. When interpreting the results one should, however, take into account that the number of observations are limited (34 research groups) and that individual deviating observations may influence the overall results (see below).

It should also be added that the measures of productivity and citations per person represent indicators that are not adjusted for field differences in publication/citation rates. Earlier studies have shown that the scientific productivity in terms of average number of articles per researcher differs between the various scientific disciplines (Kyvik, 1991, page 59). Still, a significant part of these differences can be explained by differences in the number of authors per publication. For example, on average mathematicians publish fewer articles than their colleagues in other subjects because co-authorship is less common. In our case we have therefore fractionalised the publications according to the number of authors. Nevertheless, some field differences remain, which partly may explain the low correlation found for this indicator. Similarly, although the average number of citations per person may be an appropriate indicator in intradisciplinary bibliometric studies, it needs to be field-adjusted when different scientific disciplines are compared.

On the basis of the correlation analysis there is no strong empirical argument for preferring one indicator to any of the others. Still, the indicators reflect different aspects of the research performance — aspects that may not necessarily be reflected in peer judgements. Thus, the question remains to what extent it is fair to use peer ratings as the standard. We discuss this issue further below.

The most adequate measure of the research performance is often considered to be the indicator in which citedness is compared to subfield average (relative subfield citedness). This citation index is sometimes considered as a bibliometric 'crown indicator' (Van Raan, 2000). It is therefore interesting to focus on this indicator in more depth. To give an impression of the scattering we have plotted each group in a diagram (Figure 1).

The relative subfield citedness of the groups ranged from 0.24 to 3.43. Thus, in the most extreme case one group received more than three times as many citations as 'expected' from the world average. Nineteen of the groups obtained a citation score above the world average (1.00). These groups were rated quite differently by the panels: four as excellent, eight as very good, four as good and three as fair. Table 2 summarises these results. Fifteen of the groups obtained a citation score below the world average. One of these groups was rated as excellent, 4.5 as very good, 4.5 as good and five as fair.

Although the overall correlation is not very strong, it is interesting to notice that the groups obtaining the highest relative citation indices are all rated as very good or excellent. Also the large majority of the groups being cited 150%–200% above the field average received these high ratings. A similar but less pronounced tendency was found at the other end of the citation range. Here, half of the

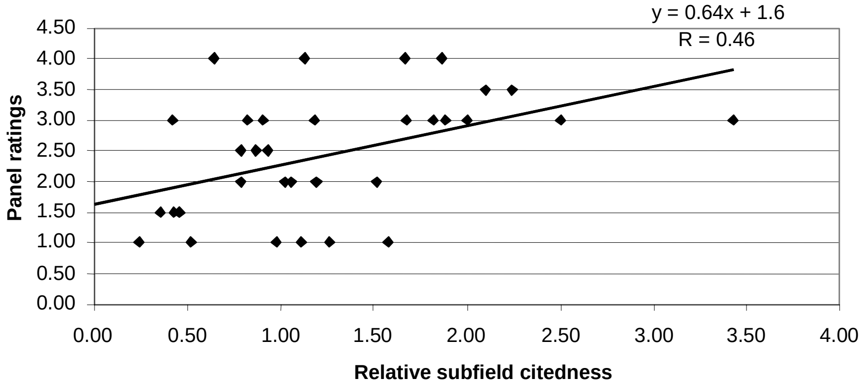


Figure 1. Graphical display of panel ratings vs. relative subfield citedness for the 34 research groups

Table 2. Panel ratings vs. relative subfield citedness for the 34 research groups, number of groups in each category

Relative subfield citedness	Panel ratings*				
	Poor	Fair	Good	Very good	Excellent
> 2.0				4	1
1.5–2.0		1	1	3	2
1.0–1.5		2	3	1	1
0.5–1.0		2	2.5	3.5	1
< 0.5		3	2	1	

Note: * Research groups receiving intermediate ratings (e.g. fair/good) have been divided accordingly.

groups with a citation score below 0.5 were rated as fair. By contrast, the intermediate interval from 0.5 to 1.5 is very heterogeneous in terms of ratings, and no positive correlation could be found. Considering the peer ratings as a standard, these findings suggest that the indicator is more accurate as a performance measure at high index values. In other words, when a research unit is very highly cited compared to the average for the field, this can be taken as a reliable indication of scientific excellence. By contrast, at moderate and low levels the indicator is more uncertain as a performance measure.

For this indicator we also checked how the results differed among the individual evaluations. We then found the following correlation factors:

- Biology/biomedicine evaluation (13 groups): 0.60
- Chemistry evaluation (8 groups): 0.52
- Earth sciences evaluation (6 groups): 0.66
- ICT evaluation (4 groups): 0.00

In the evaluation of mathematics only three groups were included, all rated as very good. The relative subfield citedness indicator for these groups ranged from 1.18 to 3.43. Thus, we identified relatively large differences in the correlation factors and a notable lack of correlation in respect to the ICT and mathematics evaluations.

As a final analysis we checked whether the correlation would be improved when using aggregated measures for each institute (calculating weighted averages). We then found a correlation factor of 0.61, suggesting that the indicator might be slightly more reliable at higher levels of aggregation.

One consequence of using the relative subfield citedness indicator is that groups producing very few publications obtain high scores if the few are cited significantly above average. By contrast, this indicator does not positively discriminate in favour of those groups having a very high productivity of papers being cited at an average overall level. In the context of research performance assessments this sometimes represents a weakness. It has been suggested that the number of citations per researcher would be a more adequate indicator when assessing research groups and institutes (Vinkler, 1998).

As described above, this indicator is still problematic when used in multidisciplinary bibliometric studies because of the field differences in citation rates. In our study we did, however, calculate a combined bibliometric indicator in which both the citation rate of the publications and the productivity of the group were taken into account. This indicator was introduced in order to simulate an assessment situation in which both the impact and productivity are emphasised.

The ‘impact-productivity factor’ was calculated by multiplying the citation index (relative to fields) with the number of papers per person (which was fractionalised for co-authorship). Interestingly, this indicator showed the highest correlation factor found in our comparative study (0.57). This is in accordance with a former study in which a combination of total papers and citations per paper obtained higher correlation with peer ratings than when the bibliometric indicators were considered individually (Anderson *et al*, 1978). Still, one might argue that such an indicator places too much emphasis on productivity to the detriment of citation impact. In any case the conclusion might be that one should not rely on a single bibliometric indicator in such contexts.

Discussion

In this study regression analyses involving bibliometric indicators and peer reviews have been undertaken. Positive and significant correlations were found for most of the indicators. However, the correlations must be considered as relatively weak. Below we examine some possible reasons why the correlation between bibliometric indicators and peer ratings was not stronger.

Limitations of peer evaluations

It is generally recognised that peer reviews have various limitations and shortcomings (Chubin and Hackett, 1990). For example, according to Nederhof (1988), the outcomes of peer reviews may be more than 50% due to chance. Given the size and scope of the Norwegian evaluations as well as the limited time available, it was obviously a great challenge for the panels to give fair assessments of the research performance at group level. As we have seen, the evaluations have also received criticism from some of the evaluated scientists. It is therefore quite likely that the evaluations have been too superficial and are misleading concerning the performance of some of the research groups.

One example is a group in marine biology rated as 'fair' by the panel but which, according to the bibliometric analysis, had a high productivity and a relative subfield citedness of 1.6. Similarly, a group in zoology rated as 'fair' published 54 articles that in total received 347 citations (relative subfield citedness: 1.12). In these cases it is likely that the panels have not given the groups genuinely fair assessment. In fact, one of the members of the zoology group officially complained to the Research Council about the assessment of his group. These examples indicate that mistaken assessment by peers is a partial reason for the weak correlation identified in our study. It also shows how biases and mistaken judgements by peers could have been adjusted if supplemented by a bibliometric analysis.

Limitations of bibliometric indicators

When it comes to the bibliometric indicators two kinds of limitations should be emphasised: methodological limitations and more general/fundamental limitations. Among the methodological limitations we will mention three examples. The first concerns the fact that only publications indexed in the ICR-database were included in the bibliometric analysis. Although the ICR coverage of the scientific production generally appeared to be good, some of the groups tend to publish more frequently in other channels than in international journals (e.g. groups doing much contract research often publish their results as reports). As we have seen, 'proceedings' papers appeared to be quite frequent in informatics. As a general guiding principle ISI-indicators are usually considered as a fair evaluation tool for those scientific subfields where publication in the serial literature represents the main channel of communication (lePair, 1995).

In our case, the poor correlation found for the ICT-evaluation, for example, could be a consequence of the emphasis given to proceedings papers in this field (among which the majority were not indexed in our ICR-database). Generally speaking, however, we think that the ICR-coverage of the departments included in our study is adequate for assessing the element of contribution to the international scientific arena and knowledge development.

A second potential methodological limitation of our study concerns the time window. As we have seen, the time period used in the bibliometric study differed slightly from those used in the evaluations. In order to see whether the results would be affected by changing the time window, we made a test. For the evaluations carried out prior to 2000 we excluded the publications from forthcoming years. However, the results showed only marginal changes in the correlation factors for the different citation indicators. This indicates that differences in the time periods considered do not significantly affect the overall results of our study.

Third, the different bibliometric indicators have particular limitations. This can be exemplified by considering a group in chemistry rated as excellent. This group had an extremely high productivity of papers and according to the bibliometric analysis the group's publications altogether received more than 400 citations. However, when measured per paper, the relative subfield citedness was only 0.65. Thus, in such cases this indicator would not be a good measure of the impact of the group. If other indicators were used instead (e.g. the number of citations per person) this group would rank much higher. This example illustrates another reason for deviations. It also shows that reliance should not be made on a single indicator when using bibliometrics in research assessment – a conclusion that has now been emphasised in a large number of studies (see e.g. Martin, 1996). This nevertheless introduces additional challenges. As soon as there is more than one indicator but the need to make clear-cut decisions still remains, the indicators must be combined into a final result. The minimum version would be a 'deliberative' combination, that is, done by a panel or a decision-maker who would discuss the importance of the scores on the various indicators and reach an overall judgement.

Concerning the use of citation indicators there are also more general limitations that can explain the weak correlation identified. For a long time there have been controversies concerning the use of citations as performance measures. Critics have argued that scientists do not cite most of their influences, citing is biased, secondary sources replace primary sources, etc. Thus, according to MacRoberts and MacRoberts (1996) and others, the assumption that citations is a valid indicator of influence has been falsified. This is not the place to discuss this general issue. However, in some cases the deviations identified in our study are probably due to limitations of citations as performance measures. For example, one group at the Department of Informatics produced 17 publications receiving only 27 citations (relative subfield citedness: 0.82). This group was nevertheless rated as 'very good'.

Limited comparability

An important question underlying this study concerns the comparability of the peer judgements and the bibliometric performance measures. Generally, the two can be expected to be positively correlated only if the aspects assessed by the peers correspond to those reflected through bibliometric indicators. To

Peer judgements and bibliometric performance measures can be expected to be positively correlated only if the aspects assessed by the peers correspond to those reflected through bibliometric indicators

the extent that they differ, the correlation will also be less (Nederhof, 1988). In our case one example is a research group at the Department of Chemistry, which was rated as 'good/very good'. However, the bibliometric analysis showed a relative subfield citedness of only 0.79. In this case the group undertook applied research of relevance for industry, etc. – a kind of importance that is not likely to be reflected through citations. Although only groups that were rated according to scientific significance and importance were selected in this study, this example shows that aspects assessed by peers do not always correspond with those measured bibliometrically. Also, panels may emphasis a more composite set of factors in their assessments. In consequence, this represents additional reasons for why we did not find a stronger correlation.

Concluding remarks

In this study of research groups at the University of Bergen we have found positive but relatively weak correlations between peer ratings and bibliometric indicators. These results may question the validity of bibliometric indicators as performance measures. However, it has been an important point of this study that peer ratings cannot generally be considered as standards to which bibliometric indicators should be expected to correspond. Instead we have found that shortcomings of peer judgements, of the bibliometric indicators, as well as lack of comparability can explain why the correlation was not stronger. This means that the level of correlation may still be regarded as reasonable and in the range of what could be expected, considering the factors discussed above.

Our results indicate that a bibliometric analysis can never function as a substitute for a peer review. However, a bibliometric analysis can counterbalance shortcomings and mistakes in peer judgements. In this way a bibliometric study should be considered as complementary to a peer evaluation. In cases where there is significant deviation between the peers' qualitative assessments and the bibliometric performance measures, the panel should investigate the reasons for these discrepancies. Then they might find that their own judgements have been mistaken or that the bibliometric indicators did not reflect the unit's performance. It has been suggested that the conclusion in such cases may depend on the type of discrepancy (Van Raan, 1996). If bibliometric indicators show a poor performance but peer judgement is positive, then the communication practices of the group involved may be such that bibliometric assessments do not work well. By contrast, if bibliometric indicators show a good performance but peer judgement is negative, then it is more likely that the peers are wrong. In accordance with these suggestions we have found that those groups in our study obtaining the highest relative subfield citedness indices were all rated as 'very good' or 'excellent'. On the other hand, the groups cited below the world average obtained rather heterogeneous ratings.

It is now frequently argued that peer review and bibliometric methods should be used in combination. This is also the conclusion of our study, since a combination of methods is likely to have improved the reliability of the Norwegian evaluations.

Acknowledgements

The authors would like to thank Arie Rip for valuable comments on an earlier draft of this paper. We also thank Ragnhild Holte-Boe for help in the collecting and sorting of data from the tenured scientific personnel.

This is an extended and revised version of a paper that was presented at the 9th International Conference on Scientometrics and Informetrics, Beijing, August 2003.

Notes

1. Some papers were published in journals that had not been assigned to a field category, or in journals for which no citation data could be obtained. These papers were not included in the bibliometric analysis.
2. This statistical test requires that the variables studied are measured using interval or ratio scales. Thus, we consider the panel ratings to be of interval type. We presuppose that the intervals between the different ratings are identical; that is, that the interval between e.g. 'poor' and 'fair' is identical to the interval between e.g 'good' and 'very good'. The Pearson's correlation coefficient found would then be independent of which numerical values we decide to give the different ratings. (For example, the result would be the same if we instead had chosen 0 for 'poor', 2 for 'fair', 4 for 'good', and so forth). It might be argued that the panel ratings should be considered as data of ordinal type. This would not allow the regression analysis. But this would not affect the main conclusions of our study.

References

R C Anderson, F Narin and P McAllister (1978), 'Publication ratings versus peer ratings of universities', *Journal of the American Society for Information Science*, 29(2), pages 91–103.

D E Chubin and E J Hackett (1990), *Peerless Science: Peer Review and U.S. Science Policy* (State University of New York Press, Albany).

J R Cole and S Cole (1973), *Social Stratification in Science* (University of Chicago Press, Chicago).

E Garfield and A Welljams-Dorof (1992), 'Of Nobel class: a citation perspective on high impact research authors', *Theoretical Medicine*, 13(2), pages 117–135.
S Kyvik (1991), *Productivity in Academia: Scientific Publishing at Norwegian Universities* (Universitetsforlaget, Oslo).
C lePair, (1995), 'Formal evaluation methods: their utility and limitations', *International Forum on Information and Documentation*, 20(4), pages 16–24.
T Luukkonen (1991), 'Citation indicators and peer review: their time-scales, criteria of evaluation, and biases', *Research Evaluation*, 1(1), pages 21–30.
M H MacRoberts and B R MacRoberts (1996), 'Problems of citation analysis', *Scientometrics*, 36(3), pages 435–444.
B R Martin, B R (1996), 'The use of multiple indicators in the assessment of basic research', *Scientometrics*, 36(3), page 343–362.
H F Moed and J G M van der Velde (1993), *Bibliometric Profiles of Academic Chemistry Research in the Netherlands* (Report CWTS-93-08, Centre for Science and Technology Studies, Leiden).
A J Nederhof (1988), 'The validity and reliability of evaluation of scholarly performance', in A F J van Raan (editor), *Handbook of Quantitative Studies of Science and Technology* (North-Holland, Amsterdam).
C Oppenheim (1997), 'The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology', *Journal of Documentation*, 53(5), pages 477–487.
Research Council of Norway (1997), *Chemistry Research at Norwegian Colleges and Universities: A Review.*
Research Council of Norway (1998), *Earth Sciences Research at Norwegian Universities and Colleges: A Review – volume 1.*
Research Council of Norway (2000a), *Physics Research at Norwegian Universities, Colleges and Research Institutes: A review – volume 1.*
Research Council of Norway (2000b), *Research in Biology and Relevant Areas of Biochemistry in Norwegian Universities, Colleges and Research Institutes: A review – panels 1–3.*
Research Council of Norway (2002a), *Research in Mathematics in Norwegian Universities and Colleges: A Review.*
Research Council of Norway (2002b), *Research in Information and Communication Technology in Norwegian Universities and Colleges: A Review.*
E J Rinia, T N van Leeuwen, H G van Vuren and A F J van Raan (1998), 'Comparative analysis of a set of bibliometric indicators and central peer review criteria: evaluation of condensed matter physics in the Netherlands', *Research Policy*, 27, pages 95–107.
A Schubert and T Braun (1986), 'Relative indicators and relational charts for comparative assessment of publication output and citation impact', *Scientometrics*, 9(5–6), pages 281–291.
A F J van Raan (1996), 'Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises', *Scientometrics*, 36(3), pages 397–420.
A F J van Raan (2000), 'The Pandora's box of citation analysis: measuring scientific excellence – the last evil?', in B Cronin and H B Atkins (editors), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (ASIS, Medford), pages 301–319.
P Vinkler (1986), 'Evaluation of some methods for the relative assessment of scientific publications', *Scientometrics*, 10(3–4), pages 157–177.
P Vinkler (1997), 'Relations of relative scientometric impact indicators: the relative publication strategy index', *Scientometrics*, 40(1), pages 163–169.
P Vinkler (1998), 'General performance indexes calculated for research institutes of the Hungarian Academy of Sciences based on scientometric indicators', *Scientometrics*, 41(1–2), pages 185–200.
VSNU (1996), *Quality Assessment of Research – Chemistry, Past Performance and Future Perspectives* (VSNU, Utrecht).
J Warner (2000), 'A critical review of the application of citation studies to the Research Assessment Exercises', *Journal of Information Science*, 26(6), pages 453–460.

8. ARTICLE VI. Citation rates and perceptions of scientific contribution

Aksnes, D.W.

This is a preprint of an article accepted for publication in the Journal of the American Society for Information Science and Technology (JASIST)

Citation Rates and Perceptions of Scientific Contribution

Dag W. Aksnes

Norwegian Institute for Studies in Research and Higher Education (NIFU STEP)

Hegdehaugsvn. 31

NO-0352 OSLO

NORWAY

Abstract

This paper presents the results of a study in which scientists were asked about their own publications and their citation counts. The study shows that overall the citation counts of the publications correspond reasonably well with the authors' own assessments of scientific contribution. Generally, citations proved to have the highest accuracy in identifying either major or minor contributions. Nevertheless, according to these judgements citations are not a reliable indicator of scientific contribution at the level of the individual article. In the construction of relative citation indicators the average citation rate of the subfield appears to be slightly more appropriate as a reference standard than the journal citation rate. The study confirms that review papers tend to be much more frequently cited than other publication types. Compared to the significance authors attach to these papers they appear to be considerably "over-cited". On the other hand, only marginal differences are found in the citation rates between empirical, methods and theoretical contributions.

Introduction

Citations indicators have been increasingly applied in the context of science policy and research evaluation. The basic assumption underlying such applications is that citations can be regarded as a measure of scientific quality or impact. During recent decades a large number of studies have been carried out to ascertain the extent to which this assumption is fulfilled. On the one hand, the question has been addressed by comparing citation indicators with evaluations based on peer-reviews; on the other, more technical issues have been examined, including topics such as database coverage, field delineation, time and field effects. Moreover, a number of sociological studies of the citation process have been carried out. Today, the dimensions of the issue are thus fairly well established, although the use of citation indicators is still surrounded by controversy (see MacRoberts & MacRoberts, 1989; Seglen, 1997; Wouters, 1999; Weingart, 2004).

The present study aims at contributing to the discussion on the use of citations as performance measures. In most of the validating studies that have been carried out, citation indicators have been shown to correlate positively with other measures of scientific impact or recognition. For example, Cole and Cole (1973) found such a correlation in respect to Nobel prizes, honorific awards and reputational characteristics. Furthermore, a number of studies have addressed the relationship between peer ratings and citation frequencies, reporting an overall positive correspondence (see, for example, Lawani & Bayer, 1983; Luukkonen, 1991; Cronin, 1996; Oppenheim, 1997; Rinia, van Leeuwen, van Vuren, & van Raan, 1998).

The question of validity can, however, be approached from different angles. When citations are used as indicators, a population or aggregate of papers is usually being analysed, representing, for example, a university department or a scientific field. One question is how citations work at these aggregated levels. This can be addressed by comparing the average citation scores of the units' publications with other performance measures such as peer ratings (see, for example, Aksnes & Taxt, 2004). Contrary to such approaches, the present study focuses on the individual publications, their citation counts and how these citation counts compare with judgements of scientific contribution. In this way the study examines the building blocks of the citation indicator. The basic idea is that an analysis of these building blocks will yield more knowledge on what the indicator stands for and represents. Thus, findings at this constitutive level will have implications for how robust and reliable the citation indicator is at aggregated levels.

In the study we have asked scientists about their own publications, focusing on the scientific contribution of the papers and their citation counts. One of the main purposes has been to assess how the citation counts of the publications correspond with the authors' own perceptions of scientific contribution. By analysing this question for different types of contributions and different types of papers the study attempts to provide a broader picture of what citations "indicate", and the limitations and biases of citation indicators.

In other words, the added value of the study is that with the help of the respondents we can look at cited papers "from the bottom up". One may then observe interesting things which may not be observed when looking from the outside in, as when the correlation between citation rates and other measures of scientific quality and impact is studied. The advantage of asking the authors is that they have first hand knowledge of the particular publications and their research fields. On the other hand, authors' perceptions may not necessarily correspond with those of other peers. Although this is an important point that needs be taken into consideration when interpreting the results, it is also the case that other peers are fallible in their judgements (e.g. biased and mistaken). Accordingly, an "objective" yardstick with which citations can be compared does not exist, and asking the authors is one point of entry to the question.

Design of study

Citation distributions are generally extremely skewed. A few publications receive most of the citations, while most publications receive very few citations or none at all (Seglen, 1992; Cronin & Overfelt, 1994; Aksnes & Sivertsen, 2004). In a previous project we focused on the high end of this citation distribution, analysing the characteristics of highly cited papers (Aksnes, 2003). The present study has been designed as a sequel to this project. As the basis for a questionnaire survey we identified the authors of the papers previously identified as highly cited. From each of these individuals' scientific production we then compiled a selection of papers to be included in a questionnaire: some highly cited, some occasionally cited and some rarely cited. By choosing scientists that had published highly cited papers we were thus able to examine papers representing the full range of the citation distributions. This also allowed particular attention to be given to highly cited papers.

The authors were first asked how they assessed the scientific contribution of the papers. In this way we could analyse how the respondents' perceptions of the scientific contribution of the publications corresponded with the citation counts. Secondly, the authors were asked about the types of scientific contributions the articles represented – empirical, theoretical, method, or review. This allowed us to analyse how different types of papers were regarded among the scientists and to what extent different citation patterns could be identified. For example, we could check whether or not method and review papers actually were more frequently cited than other papers. Thirdly, the authors were asked about the kind of contribution the publications made. Here, we developed a classification system consisting of an in-field – outside-field dimension in addition to categories for practical relevance and opening up new research avenues. A main purpose was to see how different types of contributions were assessed and cited. Fourthly, we showed respondents the citation count of the individual articles and asked to what extent these citation counts were considered to reflect the scientific contribution of the article. Based on this information we were able to address issues such as "under-citing" and "over-citing" of different types of publications.

In addition to questions on their own publications, the authors were asked some general questions concerning citations, including their general opinions on the use of citations as indicators, how a paper becomes highly cited, and on the citation history of their publications. Through these questions we wanted to assess how citations and citation indicators were perceived in the scientific community. The results on these topics will, however, be presented in another article.

Specifically, the present paper addresses the following issues: (1) How do the citation counts correspond with the authors' perception of scientific contribution? (2) Do all highly cited papers represent major contributions? (3) How do different types of relative citation indicators correspond with the authors' ratings? (4) Are review and method papers more cited than other publication types? (5) How do the authors assess different types of scientific contributions and what kind of citation patterns can be identified? (6) Do different types of papers have different temporal citation patterns? (7) Are particular types of papers "over-cited" or "under-cited" compared to their scientific contribution?

Methods

In the present study we used a total of 297 highly cited articles published by Norwegian scientists in the 15-year period 1981–1996 as an entry point (Aksnes, 2003). These papers were identified by using field specific reference standards: A paper has been considered as highly cited if the number of citations received is more than a fixed multiple of the mean citation rate of the particular subfield. (For details see Aksnes, 2003). In terms of relative citation rates these 297 papers represent the very top of a scientific production of approximately 50,000 papers published during the period.

As a first step we identified the authors of the 297 papers. For each paper we selected one author for a questionnaire survey. Because of database limitations the selection of authors was restricted to Norwegian scientists, or to scientists who had been working at Norwegian institutions for a longer time period. As a second step we identified all the papers that each scientist had authored by searching for publications in our bibliometric database. Here, particular problems were caused by the occurrence of homonymies. Because authors are indexed by surnames and first name initials only, it is not uncommon for more than one person to have identical author names. In such cases – and where possible – we selected one of the other co-authors of the paper for the questionnaire survey. This reduced the homonym problem. Nevertheless, as we will see below, some papers were wrongly identified.

For each person we selected ten papers from the period 1981–1998, including the paper originally identified as highly cited. The reason for selecting ten papers was to allow for a certain volume and spread of papers in terms of citation counts. Thus, for each author we selected one or more highly cited papers, some occasionally cited papers, and some uncited or rarely cited papers. Some of the scientists had not published this many papers during the time period. In these cases fewer than ten papers were selected. We selected articles regardless of the scientists' particular author positions of the papers.[1] Because we wanted to use a five-year citation window in the survey we did not select publications later than 1998.

In order to identify the current addresses/institutional affiliations of the authors we used a database located at our institute containing information on R&D personnel in Norway. We also searched on the Internet and used telephone catalogues etc. In total we identified 221 authors. There were different reasons for why we did not manage to identify authors for each of the 297 articles. Some of the highly cited papers were authored by the same individuals, some of the authors were dead, and in some cases it was impossible to identify a current address (one reason being change of surnames).

The questionnaires were sent out in October 2003. Among the 221 scientists selected for the survey, 166 responded (after one reminder by mail and another telephone/e-mail reminder). The response rate of 75% must be considered as quite high. Obviously, the fact that the questionnaires were “personalised” contributed to this high response rate. It is also an indication of a widespread interest in citations among the scientists.

The questionnaire consisted of three parts: one part with general questions, a second individualised part with questions concerning each of the articles selected for the survey, and a final part with additional general questions. In the second part each paper was initially presented by its title, journal name and year of publication. The authors were informed that in the questions concerning contribution they should consider the paper’s scientific importance at the time it was published and during the subsequent five years. Similarly, the calculation of citation counts was based on a five-year window (all citations, including author self-citations, were counted). The questions and alternative replies (translated into English) are presented in Table 1. In the last question (d), the authors were also asked to state the reason for their answers.

*Table 1. Questionnaire – individual publications. Questions and alternative replies**

Questions		Alternatives			
a)	How do you assess the overall importance of this paper?	Major contribution	Medium contribution	Minor contribution	
b)	Could you characterise this paper as to the kind of contribution it makes? Please select one main characteristic, even when this can be difficult.	Theoretical contribution	Empirical findings	Methods	Review
c)	Could you also characterise the paper according to the following types of contribution (select more than one characteristics if necessary)?	Uptake and immediate use within its field	Uptake and immediate use in other fields	Opening up new avenues of research in own field or other fields	Practical relevance
d)	To what extent does the paper's citation count for the first 5 years after publication (X) reflect its scientific contribution?	Large extent	Some extent	No extent	

*) In Question d) the “X” is the number of citations within a five-year window.

Various issues were taken into consideration in phrasing the questions. In the first question (a) we wanted to see how the respondents would rank the articles according to their contribution. No further definition of contribution was given because we wanted to allow for a certain flexibility in the understanding of scientific contribution. The second question (b) was introduced to allow for analyses of citation patterns of different types of contributions. The third question (c) builds on a conceptualisation of what scientific contribution consists (cf. Weinberg’s claim that what is really important is relevance to neighbouring fields (Weinberg, 1963)). The aim of the last question (d) was to examine how the respondents considered the citation counts of the articles. Here we assumed that the respondents would have sufficient knowledge to assess the extent to which the citation counts of the publications were reasonable compared to how much they had contributed to the further research (i.e. had been used by other scientists). We decided to introduce the citation counts here as part of a sentence rather than immediately following the paper title in the headline because (ideally) these counts should not influence the respondents’ assessments of the contribution of the papers (Question a).

In addition to analyses of the questionnaires, various bibliometric measures were calculated. The basis for the latter analyses was the National Citation Report (NCR) for Norway, provided by Thomson ISI. This database contains bibliometric data on individual articles for Norway (that is, publications with at least one Norwegian author address). We applied the 2003 edition of this product, with data covering 1981–2002. This database was also used to calculate the citation counts for each article which were included in the questionnaires. As described above we used a five-year citation window for the calculation of these citation rates. That is, for an article published in 1991, for example, we counted the citations to this article in the five-year period 1991–95. A five-year interval was selected because it is frequently used in bibliometric analyses and is intermediate with respect to a short-term and a long-term citation window. For some of the citation analyses, in addition to the NCR database we used the National Science Indicators (NSI) database. This database contains aggregated publication and citation counts at a national and world level for 105 different scientific fields ("de-luxe" version).

In total, the respondents examined 1578 publications. It appeared that 29 articles had been incorrectly identified because of homonymies. We were thus left with 1549 publications. In some cases one and the same article had accidentally been selected for the questionnaire for more than one author. In total 78 of the publications were examined by more than one respondent. It should be added that the number of papers behind the different analyses varies. This is because some of the respondents did not answer all the questions. Moreover, in the studies of field/disciplinary distributions double/multiple counts occur because some publications were classified by Thomson ISI in more than one category. Also a few publications were not included in the latter analyses because they had not been assigned a field category.

As an initial analysis we identified the distribution of the articles by field and publication year (see Appendix Table A1 for details). In this analysis we used Thomson ISI's *Current Contents* classification system. Here, the articles are assigned to six broad disciplines, compared to the much finer classification system used in the NSI.

Generally, the articles represent a wide range of scientific disciplines. It should be added, however, that the humanities and the social sciences were not included in the identification of the original articles and are therefore not part of this study either (except for a few papers in the behavioural sciences). By far the largest category is the life sciences (biomedicine) with 586 publications, accounting for 38% of the total. This profile is not very different from the overall publication profile of Norwegian research in the natural sciences, medicine, and technology. In terms of publication year the majority of the articles were published in more recent years. This is because the population of eligible papers was larger these years.

Calculation of relative citation indicators

Because of the large differences in citation rates between different scientific disciplines and subfields it has long been argued by bibliometricians that one should use relative indicators rather than absolute citation counts in cross-field comparisons (Schubert & Braun, 1986; Vinkler, 1986; Schubert, Glänzel, & Braun, 1988). When relative citation indicators are constructed, there are basically two reference standards that are being used: the average citation rate of the journal and the average citation rate of the field. In the first case, the citation count of a paper is compared with the average citation rate for the particular journal in the particular year. In the second case, the citation count of a paper is compared with the average citation rate for the particular field/subfield for the particular year. On this basis relative citation indexes can be calculated. The indicators constructed using these principles were originally termed the "Relative Citation Rate" (Schubert & Braun, 1986), and "Relative Subfield Citedness" (Vinkler, 1986, 1997) respectively. In this study we calculated relative citation indicators using both principles.

In order to calculate a journal specific relative citation index, for each publication we identified the accumulated number of citations using a citation window from year of publication through 2002. (Because of data limitation we could not use a five-year citation window in this analysis). We then collected information on the average citation rate of the particular journal, taking into account both the type of paper and year of publication. For example, for a letter published in a particular journal in 1995 we identified the average citation rates (1995–2002) of all the letters published by this journal in 1995. Thomson ISI refers to this average as the Expected Citation Rate (XCR) and is included as the bibliometric reference value for all publications indexed in NCR. A relative citation index (Relative Citation Rate) was then calculated by dividing the total number of citations by the XCR value:

Relative Citation Rate = Observed Citation Rate/XCR

As an example, an index value of 1.5 would mean that an article is cited 50% more frequently than "expected" for articles published in the particular journal that year.

The second indicator was calculated by dividing the citation rate of the publications by the average citation rate of the subfields in which they were published:

Relative Subfield Citedness = Observed Citation Rate/World Citation Average of Subfield

As basis we used the accumulated number of citations, adopting a citation window from year of publication through 2002. We then collected data on the corresponding average citation rate of the subfields using the NSI-database.

Other methodological issues

Different types of studies were carried out involving various univariate, bivariate and trivariate analyses. As a main issue we examined how the citation counts corresponded with the authors' own perceptions of their scientific contribution. In order to undertake correlation analyses, the authors' ratings were converted to corresponding numerical values. Spearman's rank-correlation coefficients were calculated since the requirements to calculate the Pearson's correlation coefficient were not met. It should be emphasised, however, that the correlation identified by this method is between rankings derived from the citation data and the ratings and not directly between the number of citations/citation index and the ratings themselves. Further, because the authors were shown the citation counts of each paper it cannot be assumed that the ratings and citations were totally independent measures.

The results presented in this paper thus relate to the authors' own assessments and characterisations rather than some external measure of scientific contribution. We contend that this is interesting in its own right. Nevertheless, in terms of the validation questions addressed, one should also consider whether the authors' assessments are likely to differ from assessments given by other peers. The answers to this issue will vary according to different questions in the questionnaire.

On the questions concerning type of paper (b) and type of contribution (c) the advantage of using authors' assessments is their first hand knowledge of the paper and its content.[2] Similarly, because the authors have first hand knowledge of the relevant research fields, they should be highly qualified to assess the degree of contribution (a). Here, however, "psychological mechanisms" are likely to influence the answers; in particular one might expect that the authors would tend to over-value the importance of their own works. It is an open question to what extent the authors' assessments are biased by such factors, and differ from the biases of the scientific community. Because of this and the fact that there are no final standards of scientific contribution to which citations can be compared, the results of our approach may be regarded as a gateway to the question of validation.

Concerning the question on the extent to which citation count reflects the paper's scientific contribution (d), there is the further point that there is no "correct" citation level. Since the respondents are highly cited they can be expected not to complain about being under-cited in the way an "average" scientist might do. In other words, one should expect a more balanced response. Although it may be assumed that instead respondents would tend to rationalise the reasons for frequent citations, it should be added that even among these scientists there was widespread concern about the fairness of citations. This concern was particularly manifested in their written comments.

Results and discussion

The overall results of the publication analysis are given in Table 2. Of the 1549 publications, 585 (38%) were rated by the authors as "major contributions", 678 (44%) as "medium contributions", 265

(17%) as “minor contributions”, and 21 (1%) were not rated.[3] Thus, it is interesting to notice that the proportion of “minor contributions” is by far the lowest. Apparently, the large majority of the papers selected for the survey in the authors’ views were of more than minor significance. Although one might expect that scientists in general would be reluctant to admit that they do research of minor significance, it should be recalled that the respondents are highly cited. Moreover, the papers analysed in this study do not represent an “average” sample in terms of citation rates. Considering that most papers published are poorly cited or not cited at all, our sample of papers has a bias towards the higher end.

A majority of the papers (53%) were characterised as “empirical findings”, 22% as “theoretical contributions”, 16% as methods, and 8% as review articles. We find this composition to correspond well with our expectations: empirical findings are by far the most frequent, and review papers the most infrequent. Concerning type of contribution, the majority of the publications (51%) were characterised as “Uptake and immediate use within its field”, 8% were classified as “Uptake and immediate use in other fields”, 17% as “Opening up new avenues of research in own field or other fields”, and 24% as “Practical relevance”.[4] Thus, as was expected, the largest proportion of the publications are considered to have relevance in their own field.

Concerning the question where respondents were asked to what extent the papers’ citation count reflects their scientific contribution the following results were found: “large extent”: – 53% of the publications; “some extent” – 38%; “no extent”– 9%. Thus, for the majority of the publications the citation counts were considered to correspond well with the degree of contribution. Nevertheless, there were quite a few cases where the citations counts did not correspond at all.

*Table 2. Questions concerning individual publications. Distribution, number of publications in each category**

a) How do you assess the overall importance of this paper?	Major contribution	Medium contribution	Minor contribution		Unans-wered/ ambiguous	N
	585	678	265		21	1549
b) Could you characterise this paper as to the kinds of contribution it makes?	Theoretical contribution	Empirical findings	Methods	Review		
	362	890	270	130	21	1673
c) Could you also characterise the paper according to the following types of contribution?	Uptake and immediate use within its field	Uptake and immediate use in other fields	Opening up new avenues of research in own field or other fields	Practical relevance		
	1049	163	351	483	118	2164
d) To what extent does the paper's citation count for the first 5 years after publication (x) reflect its scientific contribution?	Large extent	Some extent	No extent			
	789	569	134		57	1549

*) In Questions b and c double/multiple counts occur because some publications were classified in more than one category.

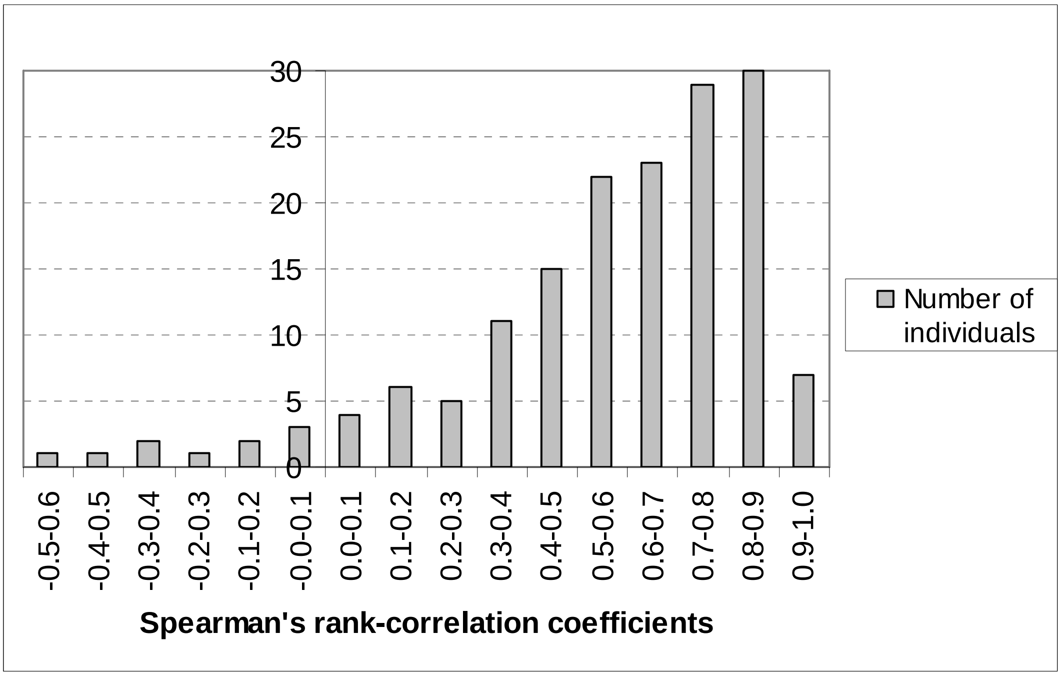
In the following we present some specific analyses and results as a basis for a further discussion.

How well do the citation counts correspond with the authors' perceptions of scientific contribution?

As one of the main issues addressed in the study we will first look at how the authors' assessments of the overall importance of their publications corresponded with the citation rate of the publications. Taking the sample as a whole, we found that the authors' ratings (Question a) correlated moderately with the citation counts (Spearman's rank-correlation coefficient = 0.49, significant at the 0.01 level). One should, however, consider that the publications analysed represented a wide range of scientific and medical research fields. It is well-established that there are large differences in the average citation rates between different subfields. On this basis one would not expect to find a very strong correlation.

What could be more interesting to analyse is the internal consistency at the level of the individual author. Here, field differences would not represent a major problem. One would expect that the publications rated as minor publications would tend to have the lowest citation counts, and vice versa. In order to assess this question we calculated the "internal" correlation coefficients (Spearman) for each of the 166 respondents. The results are shown in Figure 1 (excluding four cases with no variation).

Figure 1. Correlation analysis for individual authors. Relation between the citation rate of the publications and the authors' assessments of overall importance. Frequency distribution, number of individuals in each interval



A mean correlation coefficient of 0.56 was found at this level of individual respondents. In other words, the correlation is reasonably strong. As we can see the frequency histogram peaks in the 0.7–0.9 range. There is nevertheless a striking tail towards the left side. That is, for some of the respondents only weak positive correlations could be identified; for a few respondents we found an inverse relationship.

These findings mean that for the majority of the respondents the citation pattern reflects fairly well the internal relation of importance among the publications. But the number of respondents with a poor or weak correspondence is not negligible. For these persons, representing approximately 15% of respondents, the citation counts do not adequately reflect this internal relationship. In these cases citations are likely to be problematic as indicators of scientific contribution. We examined whether persons showing a poor correlation had any common characteristics such as particular research areas. For example one might expect that citations would be less representative as a performance measure in poorly cited fields. However, no such pattern could be identified. In a few cases one reason for the poor correlation appeared to be some frequently cited review articles that were rated as minor contributions (see below).

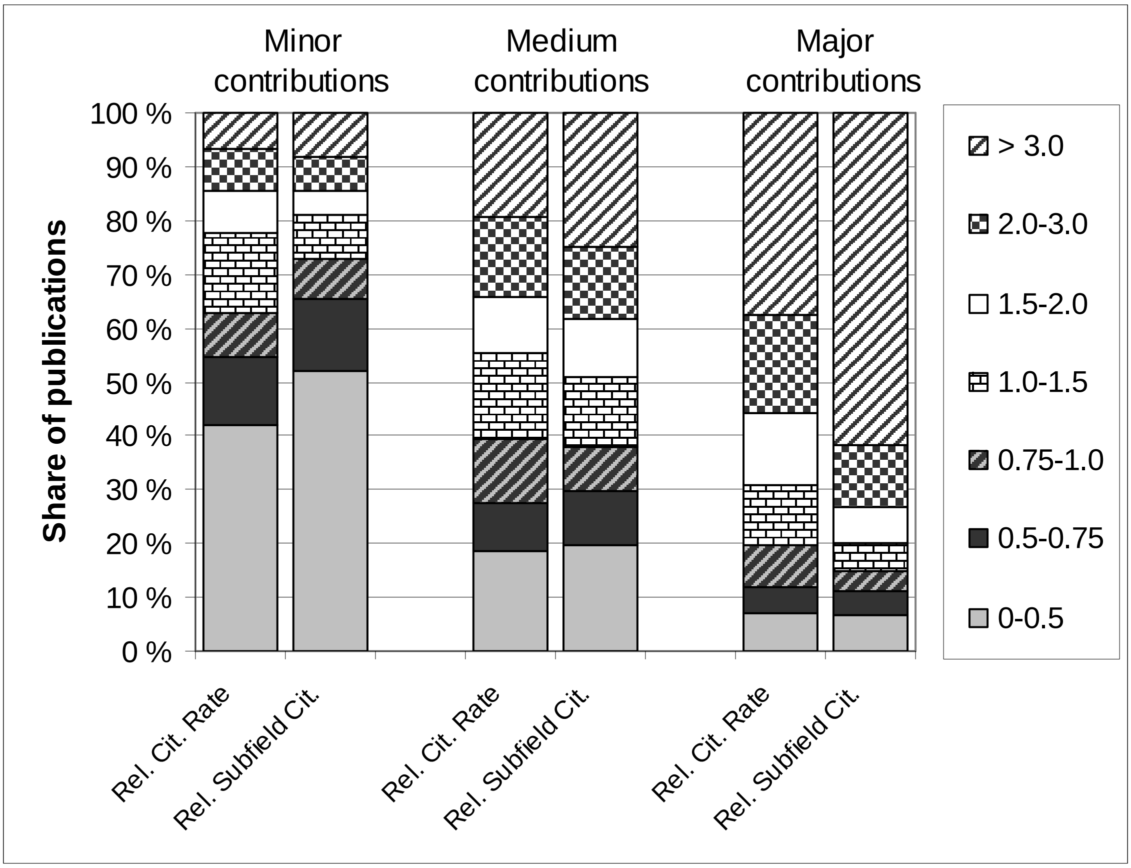
As to the validation question, how do these findings correspond with earlier studies? Although there are a large number of validation studies, it has not been usual to compare citations with author judgements (one exception being the study by Porter, Chubin & Jin, 1988, described below). However, there are several studies of the relationship between peer evaluations of papers and their citation rates. As an example, Gillmor (1975) found that the editor's ratings of articles in a geophysics journal corresponded fairly well with the citations received. A study by Virgo (1977) of journal articles in surgery and radiology showed that papers rated as important by subject experts tended to be cited more frequently than papers judged to be less important. A study of psychology journals by Gottfredson (1978) found a correlation of 0.24 between reviewer ratings of accepted articles and later citations. There are also other studies showing correlations in the range of 0.2–0.4. (For an overview of some early studies see Lawani & Bayer (1983)). On the other hand, the basis for such comparisons has been questioned by Cole (2000) who claims that the agreement among independent referees of the same articles is only moderately better than what could be expected by chance. Thus, the peer review process does not seem to be really capable of distinguishing, above the minimum level of acceptability, between articles of varying quality, importance, or validity. Compared to several of these studies, our findings appear as slightly more positive. On the other hand, the results are also disturbing in the sense that there is a subset of publications which are not picked up (or for the wrong reasons) in the citing.

Differences between relative citation indicators

We will now consider the issue in respect to the relative citation indicators. As described in the Method section we calculated two kinds of relative citation indicators: Relative Subfield Citedness and

Relative Citation Rate. The results of the analysis of how these two indicators corresponded with the authors' ratings are shown in Figure 2.

Figure 2. The relation between the authors' ratings of the overall importance of the publications and the publications' relative citation index: Relative Citation Rate versus Relative Subfield Citedness



Both types of relative citation indexes show positive correspondence with the author ratings. At the level of the whole sample, the authors' ratings correlated moderately well with the Relative Citation Rate (Spearman's rank-correlation coefficient = 0.39, significant at the 0.01 level). The analysis shows that 63% of the publications rated as minor contributions are cited below the average, while the corresponding level for publications rated as a major contribution is 20%. Furthermore, 55% of the major contribution publications are cited more than twice as frequently as the average publication (citation index > 2.0), while the corresponding share for the publications rated as minor contribution is 15%.

For the Relative Subfield Citedness an even more distinct pattern is found. Here, 73% of the publications rated as minor contributions are cited below the subfield average, while the corresponding share for the major contributions is 15%. Furthermore, 62% of the major contributions have obtained a citation index above 3, meaning they are cited three times as frequently as the field average.

Correspondingly, for this index we find a slightly higher correlation coefficient at the level of the whole sample (Spearman's rank-correlation coefficient = 0.52, significant at the 0.01 level).

Although it might be concluded that both citation indicators correspond moderately well with the authors' assessments it should be noted that the share of frequently cited papers among the minor contributions and the share of poorly cited papers among the major contributions are not negligible. Furthermore, the publications rated as medium contributions show a very heterogeneous citation index pattern. Apparently, contributions of an "average level" may have a very heterogeneous citation record, and are thus difficult to distinguish bibliometrically.

When citations are used as performance measures one or both of these indicators are usually applied. Still, there is the question of what would be the most appropriate indicator. Our results show that Relative Subfield Citedness represents the level of contribution somewhat better than the Relative Citation Rate. In consequence, the way a paper stands out in the journal is less discriminating as an indicator of contribution than how its stands out in the field. Considering the basis for calculating the two indicators this may not seem surprising. An article may, for example, obtain a very high relative score when published in a low impact journal, even if its citedness only equals the field or subfield average. Used as a measure of scientific contribution this appears as counter-intuitive. Consequently, the field average should be considered as a more adequate or fair baseline, a conclusion which is also supported by other studies (see Rinia et al., 1998; Van Raan, 2000). In the further analyses of this paper we have accordingly also used the latter reference standard when calculating relative citation indexes.

It is interesting to note that the correlation coefficient for the entire sample did not improve by using these relative indexes compared to using the absolute citation counts. In fact, a slightly lower coefficient was found when applying the journal average as a reference standard. Apparently, there are deviant citation patterns that are not resolved even when using relative citation indexes. Nevertheless, the results should not be considered as a case against the necessity of using reference standards in the construction of citation indicators in cross-disciplinary analyses. There are, in any case, strong arguments for using relative indicators in such contexts (see, for example, Schubert et al., 1988).

Differences between fields and publication years

Because the survey included a broad range of disciplines within medicine, the natural sciences and technology we also analysed whether the author ratings of contribution corresponded more strongly with the citation patterns in some disciplines more so than in others. Table 3 shows the results of this analysis for five broad disciplines using the relative citation index (Relative Subfield Citedness) as the basis for calculations. The highest correlation coefficient was found in agriculture, biology and environmental sciences (0.58), while the lowest coefficient was found in physical, chemical and earth sciences (0.49). It may be concluded that the differences at the broad level of the discipline were not particularly large, and as such there are no grounds for attaching any real importance to these differences when interpreting the overall results.

Table 3. The correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by discipline

Discipline	Agriculture, biology and environmental sciences	Life sciences	Clinical medicine	Physical, chemical and earth sciences	Engineering, computing and technology	Total
Spearman's rank-correlation coefficients	0.58	0.50	0.55	0.49	0.56	0.52
Number of publications (N)*	286	816	525	244	126	1484

*) Double/multiple counts occur because some publications were classified in more than one category. (Total based on unique publications). A correlation coefficient was not calculated for the publications in the "Other fields" category (cf. Table A1).

Another aspect of the field dimension concerns differences in the average citation rates between fields. One question is whether the correspondence between the citation level and the author ratings of contribution differ between the highly cited and the poorly cited fields. We analysed this question using the classification system consisting of 105 categories. All publications were grouped into categories according to the average citation rate of the fields they represent. The results of the analysis are shown in Table 4. The highest rank correlation coefficient (0.60) was found for the publications in the most poorly cited fields with an average field citation rate between 0 and 3, while the lowest coefficient (0.48) was found for the publications in the category above (average field citation rate between 3 and 6). Because the differences between the categories were not very large and did not show any systematic pattern we conclude that the citedness of the fields does not seem to significantly influence the degree of correspondence between the level of citation and the level of contribution.

Table 4. The correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by average citation rates of fields

Average citation rate of field*	Spearman's rank-correlation coefficient	Number of publications (N)**
0-3	0.60	76
3-6	0.48	373
6-9	0.54	504
9-15	0.55	694
15-32	0.50	418
Total	0.52	1484

*) Classified according to the world average field citation rate, based on the 105 fields of the NSI-database (reference standard: 1998 and a five-year citation window).
**) Double/multiple counts occur because some publications were classified in more than one category. (Total based on unique publications).

As an additional analysis we checked whether differences in correlation could be identified when using the age of the publications as the independent variable. In this analysis the publications were classified according to their year of publication. The results are given in Table 5. As may be seen, the rank correlation coefficients identified now show somewhat larger variation. The strongest correlation (0.71) was found for 1982 publications, while the weakest correlation (0.36) was found for 1996 publications. Although the correlation identified show large annual variation there is also a general pattern that the correlation increases with the age of the publications. However, this relation is not very strong and has accordingly no particular importance has been attached to the interpretation of the results. However, this relation is not very strong and no particular importance has been attached to these differences when interpreting the overall results.

Table 5. The correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by publication year

Publication year	Spearman's rank-correlation coefficient	Number of publications (N)	Publication year	Spearman's rank-correlation coefficient	Number of publications (N)	Publication year	Spearman's rank-correlation coefficient	Number of publications (N)
1981	0.67	25	1987	0.39	52	1993	0.43	114
1982	0.71	34	1988	0.58	61	1994	0.57	123
1983	0.61	30	1989	0.65	88	1995	0.43	111
1984	0.51	45	1990	0.63	71	1996	0.35	133
1985	0.59	55	1991	0.60	82	1997	0.54	148
1986	0.55	60	1992	0.51	102	1998	0.36	150

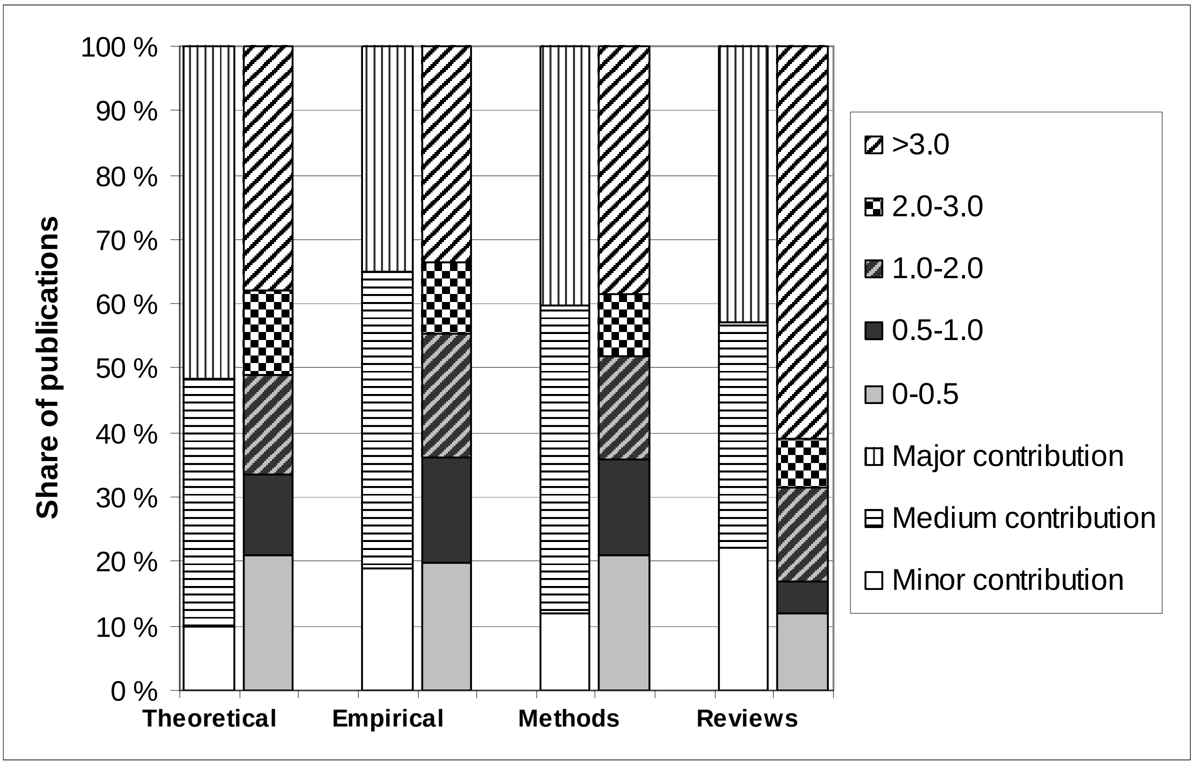
Are review and method papers more cited than other publication types?

One question frequently raised is whether different types of publications are cited more frequently than others. Several studies have shown that review papers tend to have higher citation rates that other publication types (Peters & van Raan, 1994; Moed & van Leeuwen, 1995; Moed, van Leeuwen, & Reedijk, 1996). Some have questioned the use of citations as performance indicators on the basis of such findings (see MacRoberts & MacRoberts, 1996) since the original sources are not credited when review papers are being cited.

It is also well known that some method papers are extremely highly cited. The paper on the "Lowry-method" (Lowry, Rosebrough, Farr, & Randal, 1951) (cited more than 250,000 times (Garfield, 1997)) is the prime example. On the other hand, it has it has not been empirically verified whether method papers are generally "over-cited". Garfield, for example, has argued that such papers are not generally more frequently cited (Garfield, 1979). A small study by Peritz (1983), however, showed that methodological papers in sociology journals were more frequently cited than theoretical and empirical papers. What is also less known is if there are differences between other types of papers such as empirical and theoretical contributions.

Based on the authors' own characterisation of papers, we have addressed this issue. We found the following average citation rates (within a five-year window): theoretical contributions 31 (median 14), empirical contributions 33 (median 14), methods 26 (median 13), and reviews 51 (median 24). Thus, the review articles are much more frequently cited. On the other hand there are only relative small differences between the other types of contributions. We also calculated the relative citation index (Relative Subfield Citedness) for each publication type. The results are shown in Figure 3.

Figure 3. Authors' ratings of overall importance and the relative citation index (Relative Subfield Citedness) for different types of publications.



The relative citation index corresponds well with the results above. There are only relatively small differences in the profile of the theoretical, methods, and empirical contributions. On the other hand, review articles are cited at a much higher level. Only 17% of these papers are cited below the field average (see Appendix Table A2 for more details).

Figure 3 also shows how the authors rated the scientific contribution of the publications. The theoretical papers have the highest share of major contributions (52%) while empirical papers have the lowest share (35%). Furthermore, review papers have the highest share of minor contributions (22%) and the theoretical papers the lowest (10%).

In an additional study we analysed the correlation between the author ratings and the publications' relative citation index. The results are shown in Table 6. For the theoretical, empirical and methodological contributions, the rank correlation coefficient shows only minor variations (0.51–0.57) and does not differ significantly from the overall average (0.52). For the review papers the correlation is less strong (0.42). We also carried out this analysis according to discipline (see Appendix Table A4). It appeared that for review papers the correlation was weakest in the disciplines agriculture, biology and environmental sciences, life sciences and clinical medicine (correlation coefficient 0.32–0.36).

Table 6. The correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by type of publication

	Type of publication				
	Theoretical	Empirical	Methods	Reviews	Total
Spearman's rank-correlation coefficient	0.51	0.53	0.57	0.42	0.52
Number of publications (N)*	348	869	261	117	1484

*) Double/multiple counts occur because some publications were classified in more than one category. (Total based on unique publications).

In conclusion, for the theoretical, empirical and methodological contributions the citation indicators correspond fairly well with the authors' assessments. Considering the fact that these publications have an almost identical citation profile, the theoretical contributions are somewhat underestimated in terms of citation rates (these have the highest author ratings). On the other hand, when 52% of these publications are rated as major contributions and 50% are cited more than twice as frequently as the field average, there is still fairly good correspondence.

While the study does not support the hypothesis that method papers are generally more highly cited than other publication types, it confirms that review papers tend to be much more frequently cited. It is also interesting to note that review papers have the highest share of minor contributions. This means that citations do not seem to reflect the scientific contribution of these papers adequately – the scientific value is overestimated. As mentioned above, review papers have also been considered to represent a problem to the use of citations as performance indicators because these papers only sum up existing knowledge and do not represent new or "real" scientific contributions. This problem is augmented by the fact that the review papers are highly cited. One the other hand, a significant proportion (43%) of the review papers in our survey are still rated as major contributions. This is probably because the authors have considered such contributions to have an important function in the scientific communication and publication system. Furthermore, new findings and new interpretations may emerge from the review process and can be included as part of the review article, implying that respondents have considered these papers as contributions in their own right. Consequently, this also shows that the issue concerning review papers and validity is complex. Nevertheless, it appears as a counterproductive implication that in order to become highly cited one should write review articles.

Because a review article is a separate publication category in the Thomson ISI's classification system, these papers can easily be identified (although imperfect[5]). On the basis of our findings it would not be unreasonable to suggest that if possible such papers should be excluded from citation analyses at lower aggregated levels. The problem would also be reduced if types of publications (i.e. review articles, regular articles, letters, etc.) are taken into consideration in the calculation of relative citation indicators.[6]

Do all highly cited papers represent major contributions?

During the last decade there has been an emerging interest in using highly cited papers as indicators in research assessments. One reason for this is the increasing focus on scientific excellence in science policy (Van Raan, 2000). In many countries this is exemplified by the initiative for establishing centres of excellence. In this context, highly cited papers have been regarded as potential candidates for identifying and monitoring "excellent" scientific research. Recently, this was shown in a benchmarking study by the European Commission in which highly cited papers were used as indicators for comparing the research performance of the EU countries (European Commission, 2001). Highly cited papers have also been applied as indicators in case studies of research groups (e.g. Martin & Irvine, 1983). Garfield found that high rankings by citation frequency were positively correlated with Nobel prizes in the way that nearly all Nobel laureates were highly cited within their disciplines and had produced highly cited papers (Garfield & Welljams-Dorof, 1992; Abt, 2000). Moreover, various studies on highly cited papers have been made (see, for example, Oppenheim & Renn, 1978; Moravcsik, 1988; Plomp, 1994). In the view of Small (1978) it can be assumed that frequently cited papers represent the key concepts, methods, or experiments in a field. Highly cited papers have also been viewed as ”exemplars” (using Thomas Kuhn's terminology) (Gilbert, 1977), that is the papers are cited because they represent a classical study, a ”concept” marker or show how a particular line of research is carried out.

Despite the increasing focus on highly cited papers, it nevertheless remains unclear to the extent to which these papers actually can be considered to represent contributions of major significance. Against this background we identified the most cited publication for each respondent and examined whether it was rated as a major, medium or minor contribution. It appeared that 123 papers (74%) were rated as major contributions while 43 papers (26%) were rated as medium contributions. It is interesting to compare these results with those of a study by Porter, Chubin and Jin (1988) in which various bibliometric measures were compared with the scientists' own judgements. Here, it was found that about one third of the papers nominated by authors as their best were also their most cited publication. In other words, there was no basis for considering the most cited and best paper as equivalent.

In a similar way we analysed the rating of the highly cited papers by which the authors originally had been selected for the survey (Aksnes, 2003). We recovered 266 highly cited articles, corresponding to 211 unique articles (i.e. some of the highly cited articles were assessed by more than one author). It appeared that 185 papers (70%) were rated as major contributions, 74 (28%) as medium contributions, and 4 (2%) as minor contributions.

Thus, in both cases the analyses show that the large majority of highly cited papers are considered to represent major contributions. Nevertheless, the share of "non-major" contributions is not insignificant. In other words, even papers representing the very peak of Norwegian science during a 20-year period in terms of (relative) citation rates do not unequivocally represent contributions of very large significance. This again suggests that highly cited papers cannot be considered as a simple measure or indicator of scientific excellence, at least not at the level of the individual articles. Moreover, it should be noted that there are quite a few publications the authors considered as "major contributions" that are not highly cited (see below). Our results on this issue correspond well with the conclusions of a recent study of highly cited Dutch authors (Tijssen, Visser, & van Leeuwen, 2002). Here, an author survey indicated that highly cited papers could be used as a valid measure of academic scientific excellence, but only at aggregated publication levels. At the individual level highly cited papers did not necessarily equate to a breakthrough in science or leading edge research.

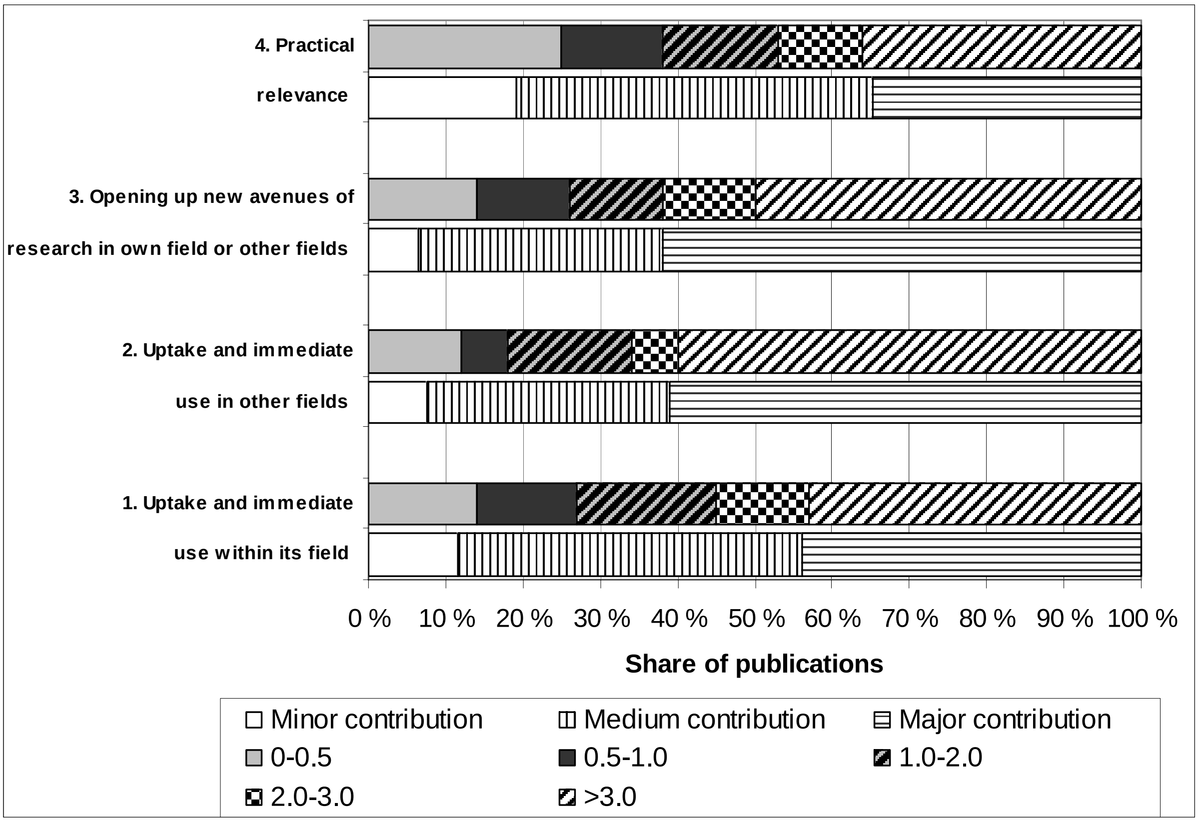
What patterns can be identified for different types of scientific contributions?

How do the authors assess different types of scientific contributions and what kind of citation patterns can be identified? Knowledge on this issue is important for a further understanding of what citations measure and their value as performance indicators.

As we have seen "Uptake and immediate use within its field" was by far the most frequent category used to classify the papers. In other words, the majority of the papers – not surprisingly – were considered to have relevance in their own field. We find, however, that when asked about the scientific contribution other types of papers were considered as more important. As shown in Figure 4, the "Other fields" and "Opening up new research avenues" publications obtained the highest average ratings. This question is linked to general ideas of what advances science. Here the latter types of contributions would be seen as particularly important (cf. Weinberg, 1963). We can conclude that respondents actually think in those terms when they have to judge the magnitude of their contribution.

In Figure 4 we have also shown the relative citation index (Relative Subfield Citedness) for the different categories. As observed, at the overall level these citation rates correspond fairly well with the assessment given by the authors: the publications characterised by making "Uptake and immediate use in other fields" are the most frequently cited, while publications with practical relevance are the most poorly cited. It is not surprising that the publications with relevance in other fields are the most frequently cited since these papers also receive external citations. On the other hand it should be noted that there are highly cited papers in all categories. It follows that papers that do not have such an external relevance may also become highly cited.

Figure 4. Authors' ratings of overall importance and the relative citation index (Relative Subfield Citedness) for different types of contribution



In an additional study we analysed the correlation between the author ratings and the publications' relative citation index. The results are given in Table 7. We did not find large variations in the correlation coefficients between the different types of contributions. For a more detailed overview of these relations we refer to Appendix Table A3 where the percentage of publications in each category has been calculated. In the Appendix more information can also be found on the correlations by discipline and type of contribution (see Table A5).

Table 7. Correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by type of contribution

	Type of contribution				
	Uptake and immediate use within its field	Uptake and immediate use in other fields	Opening up new avenues of research in own field or other fields	Practical relevance	Total
Spearman's rank-correlation coefficient	0.49	0.56	0.51	0.57	0.52
Number of publications (N)*	1020	160	345	465	1484

*) Double/multiple counts occur because some publications were classified in more than one category. (Total based on unique publications).

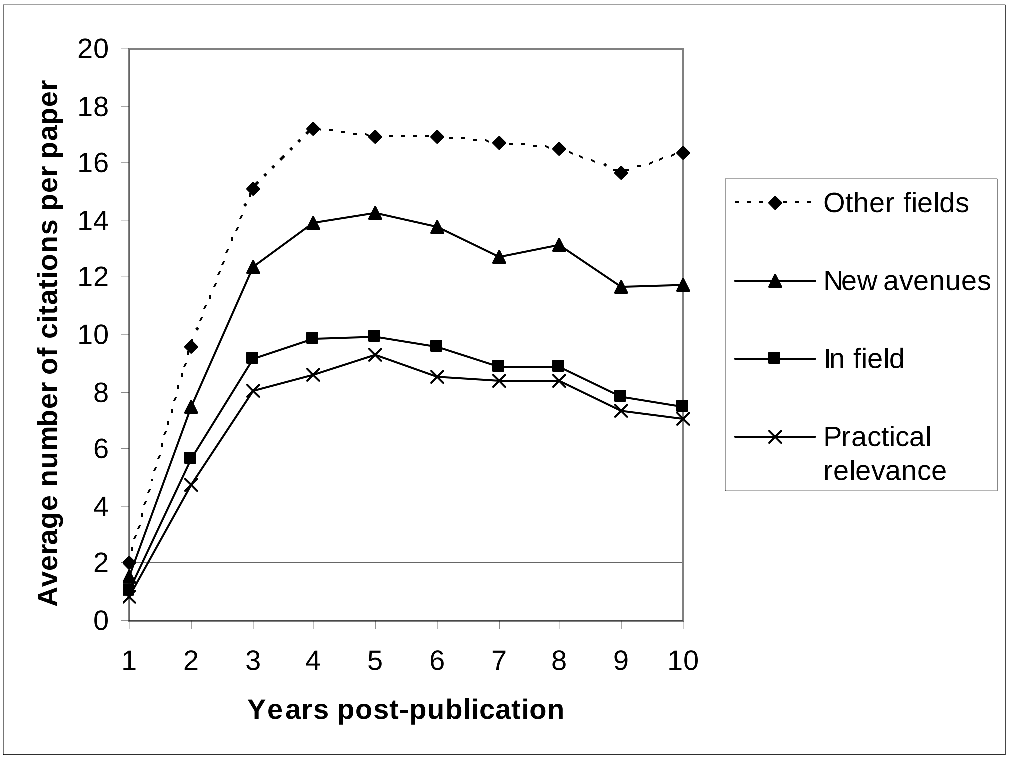
Another analysis showed that the share of theoretical contributions is highest (35%) in the category “Opening up new avenues of research in own field or other fields”. This varies between 16 and 23% in the other categories. The highest proportion of methods papers (24%) is found in the category “Practical relevance” compared with 16% in the other categories. Indeed, one would expect method papers to have more practical relevance. Thus, the findings appear internally consistent.

In conclusion the classification of papers by types of contribution shows that there are large differences in the citation rates although overall these differences correspond fairly well with the authors’ own assessment of the scientific importance of the various types of contribution. In other words, the argument that particular types of contribution will be discriminated against in citation indicators does not find support in our findings.

Temporal citation patterns

In our previous study on characteristics of highly cited papers (Aksnes, 2003) we found differences in temporal citation patterns already within the group of highly cited papers. The most marked difference was found between two types of papers: “early rise – rapid decline” and “delayed rise – little or no decline”. With our data, we are now able to check more generally whether particular types of contribution or types of paper differ in their temporal citation curves, for example involving “late rise” patterns. The results are shown in Figure 5.

Figure 5. Citations received versus time following publication. Average citation frequency per year by type of contribution.



As may be seen there are large differences in the average number of citations per year for the different types of contributions. The “other fields” publications are cited at the highest level over the entire period. Also the “new avenues” papers are very highly cited. The “in-field” and “practical relevance” papers are cited at the lowest levels.

One hypothesis we have considered is that articles relevant in other fields have a different temporal citation curve than articles relevant within the field. For example, it might take some time before a paper is “discovered” by practitioners in other fields. Consequently, one might expect that the first type of paper would have a slower start and be cited over a longer period (delayed rise – slow decline). The analysis also confirms that the “other fields” publications are declining more slowly. Nevertheless, the differences in temporal citation curves are not very large, except for the finding that these papers have visibility and are more highly cited over the period. One might expect that the patterns would be highly influenced by a few highly cited papers. However, an almost identical picture was found when plotting the median citation frequencies per year rather than the mean.

We also considered the hypothesis that articles “opening up new avenues of research” would age more slowly than other publications. As may be observed, these papers are quite frequently cited over the entire period, but the pace of decline is almost identical to the “in field” and “practical relevance” publications.

We also carried out a temporal analysis of the different publication types. We were interested in checking whether theoretical works age more slowly than empirical works and if method papers had a more long-term impact. However, apart from the fact that review papers were cited at a much higher level during the entire period no such patterns could identified.

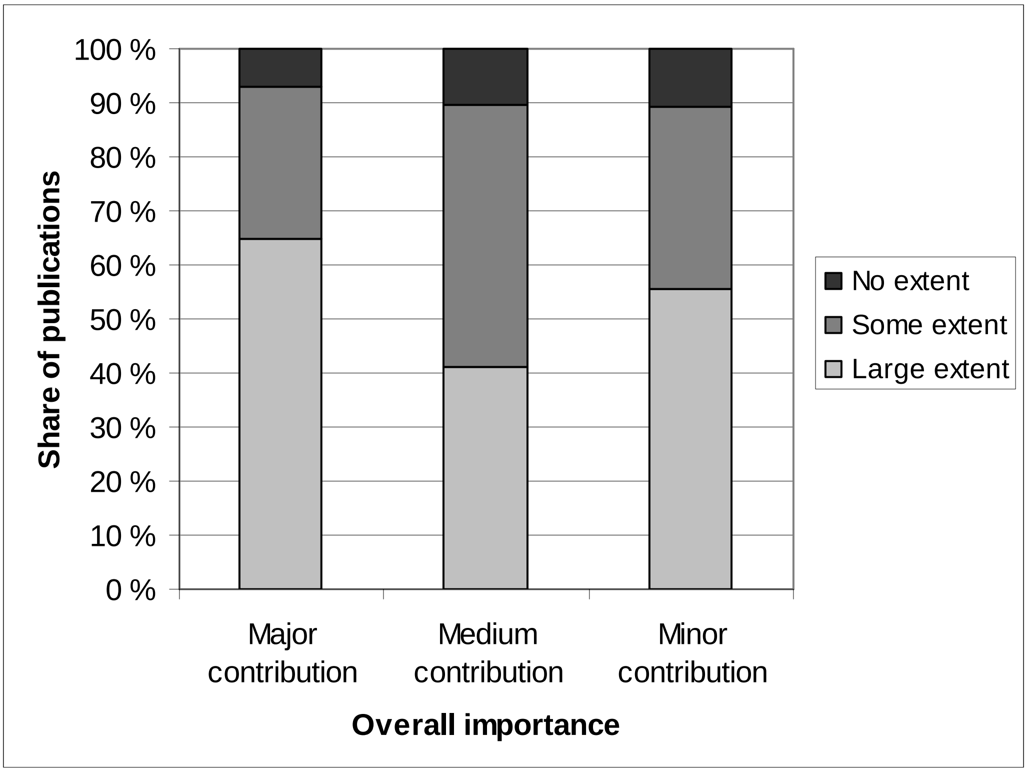
The analysis of the temporal citation curves of the different types of contribution has not revealed large differences in structural patterns. For all categories regular rise and decline patterns have been found in which the citations levels may be fairly well predicted for later years from the citation counts during the immediate years following publication. This holds for aggregated levels of publications. At the level of the individual articles there may, of course, be large variations. Accordingly, in respect of the different types of contributions analysed, an application of short-term periods in the calculation of indicators does not seem to represent a problem at aggregated levels.

Over-cited and under-cited papers

If differences between scientific quality and citation rate are random, one would expect to find similar numbers of over-cited as under-cited papers. It is difficult to check this in the absence of a fully independent measure of scientific quality. Asking authors of highly cited papers about such differences is one point of entry because they have no need to complain about overall lack of citations. In addition, we asked them for their impression of how far citations reflected the scientific contribution of their papers. This provided additional information which could be used in the analysis.

We first compared the authors' ratings of the overall importance of the papers with the opinions concerning the extent to which the citation counts reflect the scientific contribution. We found that the citation counts of the publications rated as major and minor contribution were considered to reflect the scientific contribution to a larger degree than for the medium contributions (see Figure 6). Thus, the high and low ends appear to be fairly well articulated (so that citation count and own assessment converge), while in the middle there are all sorts of contingencies at work. While this was to be expected, this is important to be emphasised because it implies an irreducible random element in using citation scores as indicators of quality. On a citation score alone, one cannot infer quality: a high score can reflect a major contribution (correlation is significant), but also contingent effects which serve to raise the score of a medium contribution.

Figure 6. Authors' assessments of the extent the publications' citation counts the first five years after publication reflect their scientific contribution, by extent of contribution



Now let us turn the question around: When are respondents seeing large, medium and no reflection of quality compared with the actual citation counts the first five years after publication? It appeared that publications with citation counts reflecting the scientific contribution were generally the most frequently cited papers. The "large extent" publications were cited 42 times (median 23) on average; the "some extent" publications were cited 23 times (median 12); while the "no extent" publications were cited 7 times (median 3). The results indicate that authors considered the poorly cited papers to be "under-cited"; in other words they consider a non-reflection of scientific contribution to occur when it is under-estimated by citations. Over-estimation is thought to occur far less frequently. This assessment must be taken seriously because it is made by scientists who have made their mark by publishing at least one highly cited paper.

In order to provide more knowledge on this issue we classified the papers according to three categories: “under-cited”, “rightly cited” and “over-cited”. We based this classification on the respondents’ answers (Questions a and d), their written comments, and the citedness of the papers compared to the field average.[7] For example, papers with citation counts considered by the respondents to reflect the scientific contribution to a large extent were classified as “rightly cited papers”. However, it was occasionally not possible to determine whether the respondents considered the papers to be over-cited or under-cited. In such cases the papers were excluded.

Based on this broad classification, some interesting patterns could be identified. First, the large majority of the papers (70%) appeared to be rightly cited. Among the rest, a much higher proportion was considered to be under-cited than over-cited (25% vs. 5%). We may conclude that there is a wide-spread opinion that quite a few publications do not get as many citations as they deserve while the contrary is much more rarely the case – even for highly cited papers. Most of the papers considered to be under-cited were rated as major contributions, whereas if over-cited papers were identified, most of these comprised the minor contributions. Second, theoretical and method papers tend to be more frequently under-cited than the other publication types, while review papers are overrepresented among the over-cited papers.

Relative to the citation counts of the publications we found the highest share of under-cited papers in the poorly cited papers, while the opposite was the case for the over-cited papers. Interestingly, almost half of the articles cited 0–5 times were under-cited. In other words, there is widespread impression that many of these papers have more importance than reflected in their citation counts. In the written comments the respondents gave several reasons for why they did not consider these low citation counts to be fair. For example, one author stated that a paper (cited just once) had been used in a court trial in the USA to stop a dangerous product. Another person said that the results of a paper (cited once) were frequently used by the oil industry in Norway. These examples also illustrate that the authors may adopt a conception of scientific contribution that not only is restricted to academic impact.

In our former study we suggested that the skewness in citation distribution is larger than the quality differentiation among scientific contributions might justify (cf. the “Matthew effect”) (Aksnes, 2003). The findings here, reflecting the authors’ own opinions, can be interpreted so as to support such a claim. In other words, if we can believe the authors, a significant proportion of the poorly cited papers should have received more citations, and thus in an ideal world the citation distribution would become less skewed. In our present, non-ideal, world we should expect visibility dynamics and rhetorical mechanisms, when citing, will continue to uphold the asymmetry between under-cited and over-cited. An immediate research management implication is that one should not take restrictive measures on scientists with low citation scores without further checks, while one can rely on high citation scores as reflecting something substantial (even though it may be exaggerated).

Conclusions

The findings of this paper have been based on Norwegian data. One question in this respect is to what extent our results may be generalized to apply to other countries. As a scientific nation Norway is rather small. Approximately 5,000 scientific articles are indexed annually by Thomson ISI. The country has research activities in a broad range of scientific specialities, but specialises in research related to its natural resources. Nevertheless, despite its small size, Norway should not be considered as a country on the scientific periphery. The country is highly integrated into the international scientific arena: most of the scientific output is published abroad and in recent years almost half of the publications show international co-authorship. Moreover, the large majority of the citations of Norwegian publications are by foreign scientists. Consequently, we do not think that the peculiarities of Norway as a research nation should be given particular emphasis when interpreting the results. On the contrary, representing a small but scientifically integrated country we think that Norway may be well suited as a case for analysing the various questions concerning citations. Of course, it is still an open question to what extent our findings have general validity. But this is a general methodological question where the most important restriction is the limited number of articles included in the survey.

In respect of the validity issues addressed in this study a crucial point is the reliability of the respondents' assessments. With our questions, we stay close to how scientists think about contributions to the advancement of science. In this way there is a *prima facie* reliability. One additional advantage of using authors' assessments is their first hand knowledge of the paper and its relevant research fields. Although there is also bias as discussed above, we consider that the results are sufficiently reliable to provide a basis for conclusions about status of citation rates, especially when used as indicators.

In the study we have found that citation counts of the publications corresponded moderately well with the authors' own assessments of scientific contribution. Generally, citations proved to have the highest accuracy in identifying either major or minor contributions. Nevertheless, at the level of the individual article citations are not, according to these judgements, a reliable indicator of a paper's scientific contribution. In particular, review articles were considered to be divergent in the way that the citation counts did not reflect the scientific contribution of these papers adequately – the scientific value was largely overestimated.

Based on the analyses we can make estimates of how much of the citation scores in a population of articles that does not reflect quality or scientific contribution. For approximately 15% of the respondents, the citation pattern did not reflect the internal relation of importance among the publications. Similarly, the deviation using relative citation indexes was in the range of 10–20% for the major and minor contributions, while publications rated as medium contributions showed a very heterogeneous citation index pattern.

Acknowledgements

The author would like to thank Arie Rip for help in the design of the project and for his valuable comments on previous drafts of this paper. The author is grateful to the many researchers who took the time to fill in the questionnaires.

References

Abt, H. A. (2000). Do important papers produce high citation counts? *Scientometrics, 48*(1), 65-70.

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation, 12*(3), 159-170.

Aksnes, D. W., & Sivertsen, G. (2004). The effect of highly cited papers on national citation indicators. *Scientometrics, 59*(2), 213-224.

Aksnes, D. W., & Taxt, R. E. (2004). Peer reviews and bibliometric indicators. A comparative study at a Norwegian university. *Research Evaluation, 13*(1), 33-41.

Cole, J. R., & Cole, S. (1973). *Social Stratification in Science*. Chicago: The University of Chicago Press.

Cole, S. (2000). The Role of Journals in the Growth of Scientific Knowledge. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Medford: ASIS.

Cronin, B. (1996). Rates of return to citation. *Journal of Documentation, 52*(2), 188-197.

Cronin, B., & Overfelt, K. (1994). Citation-Based Auditing of Academic Performance. *Journal of the American Society for Information Science, 45*(2), 61-72.

European Commission. (2001). *Key Figures 2001. Special edition. Indicators for benchmarking of national research policies*. Brussels.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Garfield, E. (1997). Dispelling a Few Common Myths about Journal Citation Impacts. *The Scientist, 11*(33), 11.

Garfield, E., & Welljams-Dorof, A. (1992). Of nobel class: A citation perspective on high impact research authors. *Theoretical Medicine, 13*(2), 117-135.

Gilbert, N. G. (1977). Referencing as Persuasion. *Social Studies of Science, 7*, 113-122.

Gillmor, C. S. (1975). Citation characteristics of the JATP literature. *Journal of Atmospheric and Terrestrial Physics, 37*, 1401-1404.

Gottfredson, S. D. (1978). Evaluating Psychological Research Reports. Dimensions, Reliability, and Correlates of Quality Judgments. *American Psychologist, 33*(10), 920-934.

Lawani, S. M., & Bayer, A. E. (1983). Validity of Citation Criteria for Assessing the Influence of Scientific Publications: New Evidence with Peer Assessment. *Journal of the American Society for Information Science, 34*(1), 59-66.

Leeuwen van, T. N., Rinia, E. J., & van Raan, A. F. J. (1996). *Bibliometric Profiles of Academic Physics Research in the Netherlands*. Leiden: Centre for Science and Technology Studies. Report CWTS 96-09.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randal, R. J. (1951). Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry, 193*, 265-275.

Luukkonen, T. (1991). Citation indicators and peer review: their time-scales, criteria of evaluation, and biases. *Research Evaluation, 1*(1), 21-30.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of Citation Analysis: A Critical Review. *Journal of the American Society for Information Science, 40*(5), 342-349.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435-444.

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*, 61-90.

Moed, H. F., & van Leeuwen, T. N. (1995). Improving the Accuracy of Institute for Scientific Information's Journal Impact Factors. *Journal of the American Society for Information Science, 46*(6), 461-467.

Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1996). A critical analysis of the journal impact factors of 'Angewandte Chemie' and 'The Journal of the American Chemical Society'. Inaccuracies in published impact factors based on overall citations only. *Scientometrics, 37*(1), 105-116.

Moravcsik, M. J. (1988). Citation Context Classification of a Citation Classic Concerning Citation Context Classification. *Social Studies of Science, 18*, 515-521.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation, 53*(5), 477-487.

Oppenheim, C., & Renn, S. P. (1978). Highly Cited Old Papers and the Reasons Why They Continue to be Cited. *Journal of the American Society for Information Science, 29*(4), 225-231.

Peritz, B. C. (1983). Are methodological papers more cited than theoretical or empirical ones? The case of sociology. *Scientometrics, 5*(4), 211-218.

Peters, H. P. F., & van Raan, A. F. J. (1994). On Determinants of Citation Scores: A Case Study in Chemical Engineering. *Journal of the American Society for Information Science, 45*(1), 39-49.

Plomp, R. (1994). The Highly Cited Papers of Professors as an Indicator of a Research Group's Scientific Performance. *Scientometrics, 29*(3), 377-393.

Porter, A. L., Chubin, D. E., & Xiao-Yin, J. (1988). Citations and scientific progress: comparing bibliometric measures with scientist judgments. *Scientometrics, 13*(3-4), 103-124.

Raan van, A. F. J. (2000). The Pandora's Box of Citation Analysis: Measuring Scientific Excellence - The Last Evil? In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield* (pp. 301-319). Medford: ASIS.

Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*, 95-107.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5-6), 281-291.

Schubert, A., Glänzel, W., & Braun, T. (1988). Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools. In A. F. J. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier.

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

Seglen, P. O. (1997). Citations and journal impact factors: questionable indicators of research quality. *Allergy, 52*(11), 1050-1056.

Small, H. G. (1978). Cited Documents as Concept Symbols. *Social Studies of Science, 8*, 327-340.

Tijssen, R. J. W., Visser, M. S., & van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381-397.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics, 10*(3-4), 157-177.

Vinkler, P. (1997). Relations of relative scientometric impact indicators: the relative publication strategy index. *Scientometrics, 40*(1), 163-169.

Virgo, J. A. (1977). A statistical procedure for evaluating the importance of scientific papers. *Library Quarterly, 47*(4), 415-430.

Weinberg, A. M. (1963). Criteria for Scientific Choice. *Minerva, 1*(2), 159-171.

Weingart, P. (2004). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics,* Forthcoming.

Wouters, P. (1999). *The Citation Culture.* Unpublished PhD-thesis, University of Amsterdam, Amsterdam.

Footnotes

1) One might argue that selecting only first author papers (i.e. papers in which the authors were first on the list) would have been more appropriate since in most cases the leading authors would be the prime candidates for being knowledgeable about the papers. However, because the selection of authors of the highly cited papers had to be restricted to Norwegian scientists (due to limitations in the available data) we could not use such a criterion (sometimes scientists from other countries were the first authors). We did not find it reasonable to include an additional first author criterion in the selection of the other papers.

2) “Secondary” contributors such as technicians and laboratory assistants, who are usually neither first nor last in the list of authors, are perhaps not likely to possess such first-hand knowledge. However, we did not identify such cases (except for one case where the respondent working as a laboratory assistant told that she could not fill in the questionnaire because of lacking knowledge). The respondents were mainly in scientific positions, quite often working as professors, although contributing in a varying degree to their different (multi-authored) papers.

3) As described above, for each author we selected publications regardless of their particular author positions. One might expect that the respondents would overrate the publications in which they have made the major contribution. We therefore checked if the first-author publications were rated differently than the publication in which the respondents were not first authors. The results showed that 477 of the publications were first-author publications. It appeared that on average the rating of these publications was almost identical as the rating of the remaining 1072 articles in which the respondents were not first-authors. Assigning a numerical value of 1.0 for minor contributions, 2.0 for medium contributions and 3.0 for major contributions, the publications on average were rated 2.39 and 2.28, respectively. Thus, no obvious “first-author” bias could be identified.

4) In this question the respondents were allowed to select more than one characteristic and 488 publications were classified in more than one category. In the analyses these publications are therefore counted more than once.

5) Of the 130 papers the authors classified as review articles, only 53 (41%) were indexed by ISI as review articles. Furthermore, 21 additional articles were indexed as review articles by ISI, but not classified as such papers by the authors.

6) The way in which the two indicators are calculated here, the publication type is taken into account in the calculation of the Relative Citation Rate (RCR) indicator, but not in the Relative Subfield Citedness indicator (cf. Method section). The reason for this is that we did not have access to the necessary data for calculating the latter indicator for different publication types. Thus, in order to reduce the “review-paper” effect, the RCR-indicator would be the most appropriate. Some produces of bibliometric indicators also calculate a field normalised citation index for different publication types (see e.g. Leeuwen van, Rinia, & van Raan, 1996).

7) In cases where the respondents’ comments could not be used to classify the paper as under-cited, rightly cited and over-cited, the following criteria (index values) were used:
Major contribution – no extent: under-cited
Major contribution – some extent: under-cited if index < 2.0
Major contribution – large extent: rightly cited
Medium contribution – no extent: under cited if index < 0.5, over cited if index > 2.0
Medium contribution – some extent: all papers excluded
Medium contribution – large extent: rightly cited
Minor contribution – no extent: over-cited if citation index > 0.5
Minor contribution – some extent: under-cited if index < 0.25, over-cited if index >2.0
Minor contribution – large extent: rightly cited
Papers with intermediate index values were excluded. Of course, it is a matter of discussion what would be the most appropriate set of criteria. Nevertheless, the intention has only been to provide a broad classification of the publications. In total 809 papers were classified as “rightly cited”, 288 as “under-cited”, and 52 as “over-cited”. 335 papers were excluded.

Statistical appendix

*Table A1. Number of publications by year and discipline**

Discipline / Publication year	Agriculture, biology and environmental sciences	Life sciences	Clinical medicine	Physical, chemical and earth sciences	Engineering, computing and technology	Other fields	Total
1981	2.5	12.5	6.5	2.5	1.0	0.0	25
1982	7.5	11.0	6.5	5.5	4.0	1.5	36
1983	2.0	15.3	8.3	3.0	1.0	0.3	30
1984	6.5	21.5	9.0	4.5	4.5	2.0	48
1985	4.0	28.0	13.5	4.5	4.0	3.0	57
1986	7.5	29.7	13.2	6.5	1.5	2.7	61
1987	11.5	22.5	7.5	8.0	3.5	1.0	54
1988	15.0	22.2	10.2	5.0	7.5	1.2	61
1989	12.5	40.8	17.8	12.5	5.0	0.3	89
1990	13.5	30.5	11.0	9.5	7.5	3.0	75
1991	13.0	43.0	17.5	6.0	2.5	3.0	85
1992	15.5	36.2	18.7	24.5	4.0	4.2	103
1993	20.3	42.0	25.7	18.3	5.0	5.7	117
1994	26.0	39.7	24.7	21.0	11.0	4.7	127
1995	15.5	46.5	21.5	17.5	5.0	16.0	122
1996	20.3	45.8	32.5	20.3	9.5	10.5	139
1997	21.0	51.7	36.7	21.5	12.0	10.2	153
1998	25.0	47.0	39.0	32.5	4.5	19.0	167
Total	239.2	585.8	319.7	223.2	93.0	88.2	1549

*) Based on ISI's *Current Contents* classification system. Articles that were classified in more than one field have been fractionalised accordingly.

Table A2. Authors' ratings of overall importance and the relative citation index (Relative Subfield Citedness) for different types of publications. Distribution of publications, per cent.

Type of publication	Overall importance	Relative citation index				
		0-0.5	0.5-1.0	1.0-2.0	2.0-3.0	> 3.0
Theoretical	Major	24 %	39 %	33 %	53 %	78 %
	Medium	43 %	50 %	60 %	44 %	21 %
	Minor	33 %	11 %	7 %	2 %	1 %
	N	72	44	55	45	132
	Overall distribution	21 %	13 %	16 %	13 %	38 %
Empirical	Major	12 %	17 %	26 %	34 %	64 %
	Medium	43 %	56 %	60 %	55 %	33 %
	Minor	45 %	27 %	14 %	11 %	3 %
	N	172	143	164	97	293
	Overall distribution	20 %	16 %	19 %	11 %	34 %
Methods	Major	9 %	13 %	34 %	44 %	72 %
	Medium	61 %	74 %	54 %	48 %	26 %
	Minor	30 %	13 %	12 %	8 %	2 %
	N	57	39	41	25	99
	Overall distribution	22 %	15 %	16 %	10 %	38 %
Reviews	Major	14 %	50 %	18 %	33 %	62 %
	Medium	21 %	0 %	76 %	56 %	27 %
	Minor	64 %	50 %	6 %	11 %	11 %
	N	14	6	17	9	71
	Overall distribution	12 %	5 %	15 %	8 %	61 %

Table A3. Authors' ratings of overall importance and the relative citation index (Relative Subfield Citedness) for different types of contributions. Distribution of publications, per cent.

Type of contribution	Overall importance	Relative citation index				
		0-0.5	0.5-1.0	1.0-2.0	2.0-3.0	> 3.0
Uptake and immediate use within its field	Major	12 %	22 %	28 %	44 %	67 %
	Medium	53 %	57 %	64 %	47 %	29 %
	Minor	35 %	20 %	8 %	9 %	3 %
	N	145	129	187	122	437
	Overall distribution	14 %	13 %	18 %	12 %	43 %
Uptake and immediate use in other fields	Major	5 %	30 %	46 %	70 %	80 %
	Medium	58 %	60 %	50 %	30 %	17 %
	Minor	37 %	10 %	4 %	0 %	3 %
	N	19	10	26	10	95
	Overall distribution	12 %	6 %	16 %	6 %	59 %
Opening up new avenues of research in own field or other fields	Major	34 %	33 %	33 %	61 %	85 %
	Medium	47 %	58 %	53 %	39 %	14 %
	Minor	19 %	10 %	14 %	0 %	1 %
	N	47	40	43	41	174
	Overall distribution	14 %	12 %	12 %	12 %	50 %
Practical relevance	Major	9 %	15 %	30 %	28 %	64 %
	Medium	44 %	60 %	55 %	68 %	32 %
	Minor	47 %	26 %	14 %	4 %	4 %
	N	117	62	69	50	167
	Overall distribution	25 %	13 %	15 %	11 %	36 %

Table A 4. Correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by discipline and type of publication

Discipline	Agriculture, biology and environmental sciences		Life sciences		Clinical medicine		Physical, chemical and earth sciences		Engineering, computing and technology	
Type of publication	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N
Theoretical	0.66	49	0.50	185	0.54	128	0.44	76	0.58	35
Empirical	0.57	177	0.52	502	0.57	320	0.45	130	0.52	44
Methods	0.61	62	0.55	106	0.69	81	0.55	46	0.66	50
Reviews	0.32	28	0.33	68	0.36	38	0.71	13	0.93	6

Table A5. Correlation between the authors' ratings of the overall importance of the publications and the publications' relative citation index (Relative Subfield Citedness) by discipline and type of contribution

Discipline	Agriculture, biology and environmental sciences		Life sciences		Clinical medicine		Physical, chemical and earth sciences		Engineering, computing and technology	
Type of contri-bution	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N	Spearman's rank-correlation coefficient	N
Uptake and immediate use within its field	0.55	217	0.47	548	0.49	302	0.53	194	0.51	93
Uptake and immediate use in other fields	0.59	27	0.61	93	0.72	40	0.35	36	0.76	8
Opening up new avenues of research in own field or other fields	0.52	56	0.49	243	0.49	135	0.46	36	0.24	19
Practical relevance	0.63	79	0.53	247	0.58	228	0.49	55	0.54	49

9. ARTICLE VII. Researchers' perceptions of citations

Aksnes, D.W. & A. Rip

To be submitted to Research Policy

Researchers' perceptions of citations

Dag W. Aksnes[1] & Arie Rip[2]

1) Norwegian Institute for Studies in Research and Higher Education (NIFU-STEP), Hegdehaugsveien 31, NO-0352 OSLO, Norway. E-mail: Dag.W.Aksnes@nifu.no
(address for correspondence).
2) The Centre for Studies of Science, Technology and Society, University of Twente, The Netherlands.

Introduction

The status of citations among scientists is ambivalent. On the one hand citations are sought-after because they are part of the reward system of science. For the scientists, being cited shows impact (through acknowledgements) and builds up reputation. On the other hand, citations are criticised for not reflecting actual scientific contribution. The use of citation indicators for research evaluation purposes puts further pressure on this ambivalence.

While issues around citations are a recurrent topic in conversations and discussions, and there is something like a shared repertoire of insights and convictions, only a few studies have investigated scientists' viewpoints on citations, and these focussed on their use for performance measures. A survey by Martin and Skea (1992) examined how various research performance measures, including citation indicators, were viewed by academics. Here, the scientist identified many limitations with using citations to assess university departments, although a large majority (67%) still felt that citations should form part of such assessments. A similar study was carried out by Collins (1991) revealing an even more critical attitude towards citation indicators. Only a few respondents in this survey gave such indicators credence in assessments of departmental research. Reasons given for rejecting citation analysis included the different referencing traditions of individual disciplines, the many non-approbationary functions of citations and the inadequacy of the databases used for counting citations. Also, the fact that the use of citation indicators often has been surrounded by controversy (for some examples see Wouters, 1999: chapter 6) may be taken as an indication of a widespread scepticism in the scientific communities towards such indicators.

On the other hand, scientists continue to use citation measures when re-affirming or questioning quality of contributions and scientific reputation. Similarly, status and A, B and C ratings of journals are sometimes linked to the citation-based journal impact factor. Thus, there are good reasons to speak of a dual attitude to citations.

We had an opportunity to investigate scientists' perceptions of citations as part of a study of highly-cited papers of Norwegian scientists. In addition to scientometric analysis of the characteristics of highly-cited papers (Aksnes, 2003a), we sent out questionnaires to scientists with at least one highly-cited paper, to inquire about the relations between the nature of their papers and the citation rates. This part of our study is published already (Aksnes, 2005). We also asked for their perceptions of citations and the role of citations, and we report our findings in this paper.

While perceptions will be coloured by the specifics of the situation of the respondents and their receptivity to the questionnaire, there are two reasons to consider our findings to reflect general features of citations in the life of scientists. First, there was no high-stakes context as would be the case when the focus is on the use of citations as performance indicators. Also, in Norway there is no use of citation indicators for funding or for science policy decisions directly affecting scientists. Second, there will be less grumbling about being always under-cited because our respondents have at least one highly-cited paper. Thus, we can expect a more balanced response.

The scientists' views may be regarded as expressions of "folk-theories" concerning citations. They are based on their experience of scientific publishing, communication, recognition and rewards, and the stories that are told about citations, rather than on systematic study. There may be "citation myths", and in some cases these could be identified by comparing the "folk theories" with findings from sociological and scientometric studies. On the other hand, valuable insights and knowledge concerning citations might be found in the "folk theories" which could be the starting point for further studies. This is particularly important because of the lacunae in our understanding of citations and their role in the world of science.

The "folk theories" circle around three issues: the relation between the quality (or importance or significance) of a paper and its citation history; the importance of visibility and how all sorts of factors play a role in determining citation in general and high citation in particular; and the fairness (or lack of fairness) of the system.

All three issues relate to the fact, recognized by scientists and documented in bibliometric studies, that citation distributions are extremely skewed. Some papers become highly cited but the large majority of publications receive very few citations or none at all. Elsewhere, we introduced the concepts of quality and visibility dynamics in order to analyze this skewness of citation distributions (Aksnes, 2003a). The quality dynamics refers to the structure of scientific knowledge and a cognitive hierarchy in which some contributions represent major scientific advances; others are filling in the details. Visibility dynamics, on the other hand, refers to certain social mechanisms that influence the citation rates, such as the bandwagon and "Matthew" effects. As a first approximation, this distinction works, and can be used to raise questions about citations. Are the citation distributions a reflection of quality dynamics and thus be "fair"? And if so, only in the aggregate, because of the contingencies around individual papers? Our respondents definitely analyse and judge what happens in terms of quality dynamics. However, they also realise that visibility dynamics is important, and are pragmatic about it. For example, some said that a poor citation count could be ascribed to the "style of presentation":

> In my view this article is "under-cited". It contains important empirical findings […] The reason is probably that the message was concealed in a boring title and there was no abstract. It is important to indicate a RESULT in the title.

There are further considerations, about the journal in which the paper is published, about timeliness, about selective citation practices. Taken together, the respondents' answers and comments offer an informal (and fragmented) sociology of citations and their role. This is what we will present and also comment upon.

The study

As respondents, we chose Norwegian scientists that had published one or more highly cited papers (Aksnes 2003a). By choosing scientists that had published such papers we could ask each of them about a range of papers, some little cited and at least one highly cited.

In total 221 scientists whose address could be recovered were selected for the survey. A selection of papers to be included in the questionnaire (maximum ten) was compiled from each respondent's scientific production. The questionnaire consisted of three parts: general questions, an individualised part with questions about each of the articles selected, and opportunities for comment plus a final question about their ideas how a paper becomes highly cited.

The four questions in the first part elicited views on citations and recognition, ending with the basic question about quality dynamics:

Table 1. Questionnaire – Part 1. Questions and alternative replies.

1. How important do you think it is for a scientist in general to have one or more highly cited papers? [Very important – Important – Some importance – Not important]
2. How important is it for you to have one or more highly cited papers? [Very important – Important – Some importance – Not important]
3. How often do you check the number of citations to your own publications? [Regularly – Occasionally – Never]
4. Do you in general think that the number of citations reflects the scientific contribution or overall importance of a paper? [Yes – It varies – No]

In addition, the authors were asked to explain the reasons for their answers to Questions 1, 2, and 4.

In Question 4, the concept of scientific contribution was used as the key term, and 'overall importance' was added to reinforce the point that papers are now seen as contributions to the knowledge reservoir (sometimes only of one research area) and help to advance it.[1] No further definition of 'scientific contribution' was given here because we wanted to allow for certain flexibility in the understanding of the concept. From the comments of the respondents we see that they actually think in terms of recognition of contributions when they consider citations.

This can be linked with the scholarly debate on citations, where the Mertonian view in which citations are considered as part of the reward system of science has often been considered to provide evaluative bibliometrics with a theoretical basis. It also leads to questions about background assumptions, especially whether scientists recognise all the works they have utilised and whether the cognitive contents of the references are all there is to referencing (Luukkonen, 1990) – which has led to studies inquiring into deviations from this ideal –, and the other assumption, often glossed over, that scientists somehow are able to recognize a contribution for what it is worth and do so immediately. In any case, it is within this tradition that concepts such as impact and recognition are used to describe what citation measures**.** We have chosen to use 'importance' in our question 4 because we wanted to keep visible possible differences between contributions to science and actual impact which might be measured by the number of citations over the years.

[1] The terms 'scientific quality' or 'a good paper' are also used, and sometimes interchangeably with 'importance'. One respondent emphasized the difference: "Uncited articles are usually of little importance, even though the quality may be good." 'Quality' refers here to the craft of scientific research (which leads to good-quality papers).

Table 2. Questionnaire – individual publications (Part 2). Questions and alternative replies.

Questions		Alternatives			
a)	How do you assess the overall importance of this paper?	Major contribution	Medium contribution	Minor contribution	
b)	Could you characterise this paper as to the kind of contribution it makes? Please select one main characteristic, even when this can be difficult.	Theoretical contribution	Empirical findings	Methods	Review
c)	Could you also characterise the paper according to the following types of contribution (select more than one characteristics if necessary)?	Uptake and immediate use within its field	Uptake and immediate use in other fields	Opening up new avenues of research in own field or other fields	Practical relevance
d)	To what extent does the paper's citation count for the first 5 years after publication (X) reflect its scientific contribution?*	Large extent	Some extent	No extent	

*) In Question d) the "X" is the number of citations within a five-year window.

The concept of contribution returns in the second part where the respondent was asked to discuss each of the selected articles. In phrasing the questions and offering alternative replies, we based ourselves on what has been discussed in the literature (e.g. special role of methods papers and review papers) and added a further aspect, originally put forward by Alvin Weinberg (1963), the difference between impact on the own field and impact on other fields. (See Table 2). Since we knew the actual citation counts we could compare them with the author's own perception of their scientific contribution. This analysis is presented in Aksnes (2005). In this paper, our interest is in the comments respondents gave, from simple ones like:

> This is the first time I have seen the citation counts – they correspond well with my expectations

or:

> Generally the citation counts are lower than one would expect and they do not always correspond well with the importance I would attach to them personally.

to more elaborate discussion:

> Sometimes I wonder why some good articles are poorly cited. Some of the articles I regard as my best publications rank at the bottom of the list – these publications have often been of indirect importance in leading to other articles that are highly cited.

The third part of the questionnaire asked for comments: a) Do you have comments on our choice of papers? b) Do you have comments concerning the citation counts/citation history of your papers in general? c) Some papers become highly cited others not. According to your opinion and experience, how do you think a paper becomes highly cited?

Among the 221 scientists selected for the survey, 166 responded. The response rate of 75% must be considered as quite high. The fact that the questionnaires were "personalised" contributed to this high response rate. It is also an indication of a widespread interest in citations among the scientists. However, not all respondents answered all the questions. While almost everyone filled in the questions where alternative replies were given, the share of respondents answering the questions requiring comments was lower, for the different questions it varied between 37% and 75%.

The respondents of the survey represented a broad range of academic fields, such as mathematics, physics, chemistry, biology, the earth sciences, clinical medicine and engineering. The majority worked as researchers within the life sciences. We note in passing that in the original selection of highly cited papers the social sciences and the humanities were excluded (for good reasons, cf. Aksnes 2003a). We did not identify any systematic differences in the views between scientists from different disciplines.[2]

Using a database located at NIFU-STEP Institute, it was possible to obtain information about the ages of the respondents. The following distribution was found: 30-40 years: 4 respondents, 40-50 years: 32 respondents, 50-60 years: 69 respondents, 60-70 years: 39 respondents, 70-80 years: 17 respondents, >80 years: 1 respondent (excluding 4 persons with missing data). The dominance of older scientists derives from the fact the original selection of scientists was based on papers published some years ago: the 15-year period 1981-1996. We will occasionally make use of the information about age, but did not generally identify any obvious or systematic difference in the views of representatives of different age groups.

We have included many quotations (translated from Norwegian to English). Quotes were selected for illustration of particular viewpoints among the respondents. In footnotes we show further, mostly similar comments.

The respondents' conceptions of scientific contribution and citation rates

An interesting background issue is already visible in the quotes above: what is to be expected as to the number of citations to an article? When we asked for each of the selected publications to what extent (and why so) the citation count reflected the degree of contribution, we were positioning the respondents as somehow knowledgeable about this. But how can respondents know what would be an appropriate citation level? Some will be acquainted with how often other papers have been cited or they might know about the citation counts of their own publications. (In the questionnaire, the respondents were provided with the citation counts of their own papers, cf. Table 2.) From their knowledge of the literature in their own field scientists can have an impression of how their publications have been referred to in other articles, and more generally, what citation level to expect.

Overall, respondents are willing to refer to something like a standard, of what is to be expected in terms of citations.[3] There are two sides to this, agreement and contestation. When respondents agree that the citation count reflects the contribution and importance of the paper, they do not need to become more specific. The standard remains implicit. When the citation count does not reflect the contribution (underciting or overciting), they contest the actual impact measure of citations and have to refer more or less explicitly to what they consider should have been the case. Most comments refer to cases of under-citation,[4] and are couched as an explanation why this could happen (as we shall illustrate later), without much consideration of what would be the "right" citation count. Comments on these issues might reflect human biases as well, for example one expects that authors tend to over-value the importance of their own contributions. Since the respondents have highly cited papers they can be expected to be more positive towards citations and not to complain about being under-cited in the way an "average" scientist might do.

A related issue concerns the respondents' view on scientific contribution. From their comments we can derive that the respondents have implicit conceptions of what a scientific contribution consists of. Quite a few of the respondents seemed to adopt a conception of contributions similar to the definition of 'importance' given by Martin and Irvine: "The 'importance' of a publication refers to its *potential* influence on surrounding research activities – that is, the influence on the advance of knowledge it would have if there were perfect communication in science (in short, if there were completely 'free market' of scientific ideas)" (Martin & Irvine, 1983). The following quotations concerning the relation between the type of journal and citation count may illustrate this: "If you publish in a 'wrong' journal you may get few citations even when the work is very good." Similarly, when commenting on the citation count of their own articles the following views were put forward:

> A 'weak' article scientifically speaking that did not elicit very much new. [It] has received more attention than it deserves because it was published in a journal with a large circle of readers.

---

[2] Due to a differing publication and citation pattern this might not have been the case if the social sciences and humanities were included.

[3] Nevertheless, quite a few of the respondents did not respond in this way. Apparently some did not feel they had enough knowledge to give such comments, exemplified by expressions such as "I have no opinion on what represent a high/low citation count."

[4] Aksnes (2005) created a composite measure for under-cited, rightly cited, and over-cited and found that 70% of the papers commented upon by respondents (about 1100 papers) appeared to be rightly cited, 25% under-cited and 5% over-cited. Review articles featured heavily in the category of over-cited.

> This is a good paper but very few […] can afford to subscribe to this journal. Papers published there are almost lost.

Comments on other issues suggest that the authors tend to adopt a conception of scientific contribution as something given which may, or may not, be recognized. For example, as reason for why one of his papers is poorly cited, one respondent argued that: “An unfortunate title, and consequently it has not received sufficient notice.”

In the work by Martin and Irvine the term ‘impact’ is used to describe a publication’s actual influence on surrounding research activities. In their views this depends partly on its importance, but is also affected by factors such as the location of the author, and the prestige, language, and availability, of the publishing journal. In other words, because there are “imperfections” in the scientific communication system, the importance of a publication may not be reflected in its impact.

Respondents tend to think of scientific contribution as somehow inherent, and visibility dynamics as a disturbance. This also implies that a paper may be seen as important, even when it has been neglected or hardly has been read by other researchers, and thus not really contributed to scientific progress.

This conception of contribution is also manifested through several comments concerning the size of the research field. A paper may represent a very important contribution, but when only a few people work on the topic, it cannot expect many citations anyway, indicated by statements such as: “This is an important contribution within a narrow field. For this the number of people that will cite it is limited.” Thus, the respondents seem to adopt a conception of scientific contribution that is independent of absolute differences in size.[5] In other words, the number of practitioners on a topic is not considered as a measure of this topic’s overall importance. Progress in small niches must be as important as progress in large research fields.

Ambivalence about the importance of receiving citations

In various ways the scientists expressed an ambivalent attitude towards citations. On the one hand the responses indicated a widespread interest in citations. Overall the results showed that the respondents were knowledgeable about citations, they had a large number of comments and viewpoints on issues such as citation distributions and visibility dynamics. However, on

[5] Another example: “It often happens that an average contribution that contains a method is cited for years and years. Other articles are aimed at a very small public and then even very good and important contributions cannot expect a high citation index.”

the question about how often they check the number of citations to their own publications, 47% answered “Never”, 49% answered “Occasionally”, and 4% “Regularly” (n=165). Considering the fact that the authors have been highly cited it might also be considered as surprisingly that almost half of the respondents never check the citation rates of their publications.[6] In this respect, the use of bibliometric indicators for science does not seem to have much impact yet (cf. Weingart, 2005).

The large majority of the respondents considered it as either important or very important to have one or more highly cited papers, cf. Table 3.

Table 3. Researchers’ opinions on the importance of having highly cited papers.

	Very important	Important	Some importance	Not important	Total (n=166)
How important do you think it is for a scientist in general to have one or more highly cited papers?	52%	41 %	6%	1%	100%
How important is it for you to have one or more highly cited papers?	45%	37%	13%	5%	100%

The slight but definite difference between importance in general and importance for oneself can be a matter of projection: “others run with this, but I am more realistic”. A few of the respondents noted that they might not be typical, because they had their established reputation and need not go for highly-cited papers anymore.[7]

Various reasons were given by the respondents for why it was important to have highly cited papers. A frequently mentioned reason was that it would prove the scientific impact and quality of one’s work, as the following quotation illustrates: “It means that the article is a valuable contribution to the field and is being used when others write/do research within the same field.”[8] Immediately linked to this is the consideration that quality brings reputation (cf. quality dynamics). Whatever the link between quality and citations, since reputation is important, citations can be sought in their own right. To further their career

[6] One might expect that the youngest scientists would be more familiar with citation indexes and thus check their citation counts more regularly (e.g. because of higher adaptation to new science policy regimes, familiarities with web-based search technologies). The average age of the respondents that never check their citation counts and the average age of the respondents that occasionally check their citation counts were, however, almost identical (57 and 56 years, respectively). Thus, no such systematic pattern could be identified. Even scientists in their seventies (retired) occasionally check the number of citations.

[7] This can be linked with Mulkay’s (1972) analysis of risk taking in science: the young have little reputation to lose, so can take risks. The old, or better, the highly reputed scientists, can permit themselves to risk their reputational capital (Linus Pauling would be an example). In between, reputational capital has to conserved and increased – so highly-cited papers are necessary.

[8] Other examples: “It shows that the work has (usually) been of high quality / has been useful / has been of importance for others research / has represented a hallmark or been of high international standard / to a certain degree reflects that the articles are read and that your research is of interest to others.”

and/or obtain credit to mobilise resources, scientists will apply themselves to publish and improve their chances of getting citations. And when these do not materialize, find excuses related to visibility dynamics leading to discrepancies between citation counts and (self-assessed) contribution to science.

Most of the respondents focused on issues related to the personal sphere, that is, why it is important for the researchers themselves. Here, the importance for scientific recognition and for scientific career was often mentioned, but some also emphasised personal motivation. The following quotations taken from different respondents' questionnaires illustrate these various types of issues:

> Gives you recognition in respect of colleagues, and a scientific reputation internationally
> Highly cited papers are important in order to get scientific positions / are important for getting access to scientific collaboration
> as an employee in private sector selling know-how, this means that nobody questions your scientific sincerity
> important in order to get money for further research / used by funding agencies and when applying for positions (in some cases)
> give the researchers legitimate self-confidence / self-esteem, reduces faint-heartedness/depression.[9]

From these comments we see that the respondents attach large importance to highly cited papers. Citations are sought after; they are valuable because they in the view of respondents increase recognition and make funding easier.[10]

Clearly, there is something at stake in receiving citations. Thus, it becomes important to explain discrepancies between actual citation counts and importance/contribution. Explicit and implicit criticism of citation counts is stimulated because of this. What happens also is insistence on contingency, up to randomness – "A paper is often 'discovered' more or less by chance", "It is coincidental what is cited". At the same time, and for the same reasons, citations are upheld as an important indicator of quality and achievement. Thus, a mixture of criticism and agreement.

---

[9] Other examples: "gives one a world-wide reputation as a scientist / highly cited papers are important in order to obtain scientific positions / helps your career / leads to invitations to join various research groups internationally / important for salary increases / important in evaluations / important for the institute's reputation / highly cited papers pay off in terms of funding of project applications / highly cited papers may be a necessary condition for obtaining funding / is an inspiration for further research / shows that you have been successful."

[10] Would they want reputation without quality? Some do, an extreme case being fraud, as has happened especially in life sciences (Broad & Wade, 1982).

Discrepancies between citations and scientific contribution

As reported in Aksnes (2005), the citation counts of the publications selected for the questionnaire were found to correspond reasonably well with the authors' own assessments of their scientific contribution, but only at the aggregate level.[11] At the level of the individual article citations are not, according to these assessments often argued in detail, a reliable indicator of a paper's scientific contribution. Illustrative is this quote:

> Some of the articles which I regard as my best papers are poorly cited. An example: a paper from 1980 showed that […] contrary to what had been found by other highly recognised scientists. Our findings were ignored for a long time (and not cited), despite being published in a very good journal. Eight years later other studies showed that our results had been correct. The article was then cited for a short period, but a little later other "follow-up" articles were cited instead of our article. Another example: a Nature article represented a major contribution within the field and has been cited approx. 100 times. I would have expected it to have been cited more frequently, but instead most scientists cite follow-up studies published some years later.

When asking the authors to comment on the citation counts of their own publications quite a few remarked that these did not adequately reflect contribution:

> The citation counts should generally have been higher.
> Too many articles are poorly cited or not cited at all.
> A relatively normal citation history. Articles I regard as good have been neglected, other more or less unexpected observations have "caught on".

There were also respondents that regarded the citation counts of their publication to correspond well with their assessed contribution:

> Interesting patterns which strengthen my opinion that they give a representative picture of quality, at least within my field.
> The number of citations usually reflects the articles' scientific contribution, but not always.

The analysis in Aksnes (2005) showed that generally citations best reflected major or minor contributions, while in between there was much variety, as some respondents recognized:

> Uncited articles are usually of little importance, even though the quality may be good. Articles that are medium cited can hardly be separated significantly – five or fifteen citations do not matter.

When commenting upon their own articles, the respondents mainly focused on the cases were the citation counts were misleading – because then there is something at stake:

> The article is much more important that the citation rate suggests.

---

[11] In 53% of the cases the respondents considered the citation counts of the publications to reflect the scientific contribution to a large extent, while 38% were considered to reflect the contribution to some extent. In 9% of the cases the citation counts reflected the contribution to no extent (cf. Aksnes, 2005).

> I regard this paper [cited 10 times] as one of my most important publications.
> This must be a mistake. This article is frequently cited and one of the most important I ever have written [uncited].
> Quite frankly, this article is too poorly cited. After all, it laid the foundation for the articles x and y which have been massively cited.

The respondents also draw attention to particular circumstances where citation counts would be misleading, for example:

> Some articles that contain wrong conclusions are frequently cited because other scientists then will cite the work and show that they have new and correct results.
> If an article [in mathematics] completely solves a problem, it will sometimes be poorly cited – then there are no related problems to work on.

The respondents gave different arguments for why they considered the citation counts to be unfair (and unfairness tends to refer to cases of neglect, the uncited and poorly cited papers, not to overcited papers). The quotations below exemplify the range of these comments:

> The content [of the article] was used in a court trial in the USA to stop a dangerous product [from being marketed] (the product was stopped) [cited 1 time].
> A good article with large relevance for clinical practice, it is surprising that it is not cited.
> A paper of local character, but often applied by the oil industry [cited 2 times].
> The citation count [0] surprises me, I have received many inquires for off prints.
> The topic is relatively "narrow" and only a few people will use the results. However, the number of indirect users is many – but these do not result in citations.
> We showed that a well-known and highly cited work of German chemists was WRONG. It terminated a discussion that nobody saw a reason to continue, not even those who published the original work (we have, however, received positive verbal feedback) [uncited]. [12]

There were also several cases where the authors considered it as fair that some of their articles were poorly cited:

> This article contains little new compared to article x, and I see hardly any reason why it should be cited.
> The article confirmed that some previous data in a particular research area were "correct" and in this respect it did not have much novelty.
> It showed a new and useful use of a medicament, but represented a minor study compared to the [similar] other studies.
> This is an investigation of Norwegian conditions and of little general interest in the wide world.
> The work is mainly a descriptive survey of a particular geographical area [uncited]. [13]

---

[12] Other examples: "Later works that have used ideas from this article have not cited it because it was considered as unimportant at the time it was published. / We received many inquires about this method in the form of letters, and it has therefore aroused interest. It is a bit strange that it is cited so infrequently [1 time] ….. / The lack of citations means that the work mainly had a national importance by forming the basis for a White paper on cancer. / This article has later had a relatively large importance for the understanding of otherwise difficult rocks [uncited]. / Shortly after publication a book was published in which these results were included. The book is frequently cited approximately 300 times [uncited]. / This is basic new biology that has become text book material [cited 10 times]. The field is poorly cited. / This is a very uncommon illness and cannot expect many citations. However, scientifically it is important and the citation count should have been higher."

[13] Other examples: "This was a study of low quality, with poor data and a negative result that was difficult to interpret. / A product that never was released on the market, mainly because of poor effect in terms of blood

Publications with citation counts considered by the respondents to reflect the scientific contribution to a large extent were the most frequently cited papers. When comments were given on these papers aspects related to scientific quality and significance were often used to explain why they had been much cited:[14]

> The article is highly cited because it opened up a new research area.
> The first clinical test showing the effect of an epoch-making medicament.
> The article tries to examine the degree of evidence of the effect of a new treatment that had received much attention. The result was negative. That is, the treatment did not have any effect. The article has been prize-winning.
> This article was ground-breaking and hardly any other article has been more cited.

When general views were asked for, aspects related to scientific significance and quality were similarly often used as explanation for frequent citations:

> Important contributions either methodologically or epistemologically will necessarily become highly cited.
> Articles that open a new research field or change the direction of a present field are rewarded with many citations. The first good articles in a field will usually be remembered longest and these articles are therefore most frequently cited over time.
> [Highly cited when] it has contributed to a paradigm shift or lead to new ways of thinking in the field.
> High citation counts can indicate high quality/value but also high actuality.
> A paper is highly cited if it describes new knowledge that can be used by other researchers.
> Results that remain standing [are highly cited].[15]

Clearly, these are folk theories about citations, and sometimes convictions about how the (reward) system should work. At the same time, they do reflect experiences, even if these are mixed. This combination explains why frequent citation is more often considered as to reflect scientific importance – although cases of “over-citation” also are reported.

The folk theories can be quite sophisticated, for example when particular types of research are seen as being favoured in terms of citations – while the resulting counts do not

---

pressure reduction/ When we submitted this paper for publication we knew that it was not important, but we could not know this when we commenced the research.”

[14] One might expect that the respondents would tend to emphasise quality when explaining why one of their papers has become highly cited and visibility dynamics as reasons for poorly cited papers. We could, however, not identify such a tendency and there were also many contrary cases, i.e. where low quality and little significance were used as explanation for poor citation and frequent citations were explained by visibility dynamics. Also one might expect that there would be a contrast between the views the respondents have in general and their views on their own situation. To a certain extent we did find support for such differences. Respondents list a large number of reasons why citation counts do not reflect contribution, but in the case of their own articles there is overall still a fairly good correspondence between the citation counts and their assessed contribution. At the same time it is difficult to identify any obvious antagonism between their general views and the comments the give on their own articles. In both cases a variety of issues are addressed.

[15]

necessarily correspond with the papers' scientific contributions, as the following views on highly cited papers illustrate:

> Large clinical experiments are most frequently cited
> [Highly cited when] crossing field boundaries
> Highly cited are particularly new methods, new and important empirical findings, and results of new treatments and investigation methods.
> [Papers on] a new illness are much cited.
> Clinical relevance gives many citations.
> The contribution must be of a type that other can build upon.
> [High citation counts]: reference articles (standard values).

Further nuances are visible in how evolution over time, the temporal dimension, was recognized. For example, the following arguments were given for why some papers become highly cited:

> A paper will be highly cited if it is the first to arrive with new data upon which many other build their further research.
> Articles are frequently cited if they contain good research and address an issue that is of current interest.
> The topical interest at time of publication is more decisive than the scientific value of the article.

But the temporal dimension was also used as argument for why the number of citation not always reflects scientific contribution, for example:

> If you are ahead of your time you will not be cited.
> If your work is something entirely new and significant it may be a long time before it gets cited.
> A work may appear as very original and important for some time, but is then rejected by subsequent research.

This dimension was also addressed in the comments on their own articles. For example, the respondents sometimes used expressions like "published at the right time" for frequently cited papers and "rapidly superseded by later works" for poorly cited papers:

> This contribution was published too late for the debate, that is, the issue was not anymore in fashion.
> The article was written at the right time and therefore obtained a particularly high citation rate.
> It hit the bull's eye in respect of what was the current vogue.
> Important when it was published, but not anymore.
> An unexpected result that probably will attract more attention later.
> Summing up […]. The subsequent development has been very rapid and citing is less topical.

For some of the articles (usually the poorly cited) there were also comments concerning "delayed recognition":

This approach was not in common use until 10 years later. / The work is important for the understanding of blood circulation. It is not yet ‘discovered’. / An important method that did not become popular until 8-10 years later.

Phenomena such as delayed recognition were considered to affect the validity of citations as measure of scientific contribution. The respondents here expressed what has been a common argument against citation as indicators. However, it has been shown statistically that delayed recognition as reflected through citation patterns is generally very rare (Glänzel, Schlemmer, & Thijs, 2003). Thus, on this point the respondents seem to rephrase what might be considered as a citation myth among scientists although a few of them did have real-life experiences with publications that had been recognised many years after publication.

Visibility dynamics and obliteration

In their comments the respondents directed attention towards various mechanisms disturbing a simple correlation between contribution and citation count. In the citation process visibility and recognition were argued to play important roles:

It is important that an article starts to get cited so that other scientists become aware of it.
There is a bandwagon effect: citations lead to new citations.
Citation involves a chain reaction, people are following each other.
Key publications are often cited at a disproportionate rate.

Others pointed to social hierarchy as important:

Citations depend more on the reputation of the scientist more so than the significance of the article.
Fewer people read the article when the author is not an established scientist.
Within my field I notice that citation is a matter of fashion, some authors (big shots) appear very frequently, despite the fact that not all of what is being cited is that good. One will be frequently cited if one has made a name for oneself.

Problems related to self-citations and what was considered to be a social bias in the citation process were also mentioned by a few of the respondents:

Citation is often a “clique”- affair, they cite their own work, while other (uncited) works have equal or more relevance.
You have “friend” citations within some groups.

Furthermore, when commenting the citation counts of their own articles some of the respondents considered international collaboration as important:

It is frequently cited because the publication was authored by a large number of authors from different countries.
This [article] is a description of an experiment with many participants [authors], which means that many of them will cite it subsequently

Others indicated that there are geographical biases involved:

> American scientists seldom cite European scientists within our field.
> American scientists rarely cite articles from European journals.

According to a few of the respondents the title and abstract and of a paper and the style of presentation is more important for citation rates than the quality and relevance. The following comments concerning why papers become highly cited exemplify this view:

> A sexy title and an abstract that promises more than the article contains is the safest way to high citation frequency.
> The description must not bee too complicated [in order to get highly cited]
> It is important to "sell the article". Title and abstract are extremely important in order to catch on.
> The results must be clear or definite (not: on the one hand… on the other hand.).

Others claimed that there is a considerable element of randomness in the citation process:

> A paper is often 'discovered' more or less by chance.
> It is coincidental what is cited.
> It is just a matter of luck. Citation rates are relatively misleading.
> Because of limited space you need to make a selection of articles to be cited. Many relevant works are being omitted. Citations are based on, for example, knowledge of a researcher/group contributing to the field (here there is a major weakness)…

In their comments on such mechanisms, respondents may well be reproducing a shared repertoire among scientists about citation processes, up to what might be considered as citation myths. There is definitely evidence, as when respondents describe their experience:

> This article was ground-breaking and hardly any other article has been more cited. However, it may be an example of 'over-citing' because of the strong commercial implications of the article.
> Competitors in this field practice citation collaboration where Europeans are not wanted. A number of articles published later should have cited this article.
> The article clashed with the opinions of some of the leading international research groups within our area. Their reaction was to 'boycott' the article which should have received higher citation counts."

The question, of course, is how generalizable such experiences are. Scientists feel sufficiently certain about their folk theories to transform them into advice.[16]

In passing we note that the respondents tend to mention issues that are also the focus of discussions of the use of citations indicators. For instance, Seglen (1997) mentions "friend citations" and bias towards US science as examples of problems associated with the use of

[16] Examples in the comments of respondents for the particular case of publishing in conference proceedings: " One should tell the PhD students that it is a complete wasted of effort to publish in conference proceedings."

citation data. Similarly, bandwagon and “Matthew” effects have been recognised as mechanisms operating in science since the pioneering studies by Merton and Price (Merton, 1968; Price, 1976). There is however, little knowledge on the actual importance of such mechanisms on the distribution of citations. Garfield has argued that citation circles are much talked about, but rarely, if ever, documented and identified (Garfield & Welljams-Dorof, 1992). Other issues could be documented trough bibliometric studies, for example, multi-authored papers have been found generally to be more highly cited than single-author papers (Aksnes, 2003b) and internationally co-authored papers have been shown to be more cited than single country papers (Narin, Stevens, & Witlow, 1991).

The complement of visibility dynamics is obliteration dynamics: when one paper becomes more visible, others will become less so. Various mechanisms are mentioned in general:

> Sometimes an original article is poorly cited, while a secondary article based on the original article obtains all the citations and honour.
> A very important article opening up new avenues of research is not necessarily highly cited because following-up articles gradually become cited instead of the original article.
> The citation count of an article depends on if/when it is included in a review article: if a review article is published soon after publication this article will be cited instead the original article.

Also when commenting upon the citation counts of their own papers, obliteration is mentioned, often related to competition and timeliness:

> The scientific quality of the article is very good but a competing article had been published shortly before in NEJM [The New England Journal of Medicine].
> There are some very interesting findings reported in this article but other research groups had previously published similar results in very high impact journals, and these articles are usually (and rightly) cited instead of our article.
> A progress report. Sometimes it is necessary to publish quickly in order to avoid that being forestalled by competitors. It is reasonable that this article is not much cited.
> One article among an enormous number of publications concerning this topic, it drowns in the “crowd”.
> The article is a part of a serial. Each article contains a limited amount of new information, and one would not expect each article to be a bestseller.
> This article is identical with article x. The differences in citation rates reflect the different circle of readers. [17]

“Even good works in conference proceedings are poorly cited. If I had started my scientific career today I would have adopted another publishing strategy: not to publish in books and conference volumes.”

[17] Other examples: “The article was important when it was published but a few years later it was replaced by other more detailed investigations. / What the article describes is ‘good to know’, a bit boring and no reason to cite it (once one knows it). / Interesting findings were reported in this article but we continued the work in later papers and these are more highly cited. This explains the low citation rate of the article. / At the same time other research groups published similar findings. Which articles that are being cited therefore vary. Furthermore, the ‘follow-up articles’ are now being cited to a larger extent. / This article is only a retelling of the work presented in article x.”

Citing of review articles rather than the original articles has often been considered as a problem for the use of citation indicators because the original sources will then not be credited (MacRoberts & MacRoberts, 1996; Seglen, 1997). The respondents are well aware of this problem and how it affects the citation counts of the individual research papers in different ways. The respondents also point at similar biases appearing when follow-up articles are cited instead of the original articles.

The phenomenon of "obliteration by incorporation", meaning that when a work is sufficiently well known, it is not cited anymore, was recognized by one of the respondents: "This [highly cited article] has become a standard technique for serological (…) It is now standard so that it is not cited anymore." The phenomenon has been considered as a problem for citation analysis, although Garfield has argued that: "There is, however, not much chance of obliteration causing inequities. It happens only to work that makes a very fundamental and important contribution to the field; and before the obliteration takes place, both the citation count and reputation of the scientist responsible for the work usually reach a level that makes additional credits superfluous" (Garfield, 1979). Garfield's response is too limited however, given the variety of obliteration effects when follow-up publications eclipse the original sources, as testified in the experiences of our respondents.

Effects on citation counts

Several respondents considered the type of publication as influencing the citation rates:

> Citation rates depend on the types of paper. Methodological papers and review papers are often frequently cited.
> Review articles that are published by persons who have already published much within the field are often frequently cited.
> New methods are often cited if they are good and useful. In addition to review articles the most frequently cited papers are key methods.
> Sometimes the citation count is high because the article describes a method that many scientists can use, although the scientific importance is not that much.
> Theoretical works need to have relevance for experiments in order to get cited.

These are general viewpoints, but they were supported when the respondents commented upon their own articles. Review articles drew most of the comments:

> This is a review article. For such publications the citation frequency is highly influenced by where the article is published, the number of research groups working within the field, and the number and quality of other corresponding review articles.

A review, cited at a disproportionate high rate compared to article x which is the original article. [18]

Here the respondents' views are in accordance with bibliometric findings. Several studies have shown that review papers tend to have higher citation rates that other publication types (see Moed & van Leeuwen, 1995; Moed, van Leeuwen, & Reedijk, 1996; Peters & van Raan, 1994). This was also the case in the set of papers used for the questionnaire (Aksnes, 2005).

A very frequent view was that citation counts imply a bias against small and specialised subfields:

Citation counts reflect activity (or lack of activity) within the field.
High citation counts may be a result of many scientists working on the topic, quality may not be reflected.
If the field is large a central article will be cited by many.
If you work in a small field you will pay the price.

Similar arguments were also used when the respondents commented on the citation counts of their own articles:

[The article is poorly cited] because there are not many scientists who use the experimental technique described in the paper.
This is a scientific niche with little publishing activity. Because of this the chances of getting cited is limited.

Some also emphasised that the traditions for citing differ between fields, causing inequalities when comparing publications in different fields. This argument was, however, much less frequently mentioned that the size effect.

The idea that working in small research fields is a disadvantage in terms of citations is a misconception, at least it is not supported by bibliometric findings (see Peters & van Raan, 1994). Garfield, for example, has emphasised that a large field produces many references, but there are also more publications that need to share those citations (Garfield, 1998). Thus, the average citation rate per paper is not influenced by the size of the field. What is still the case is that the size of the literature affects the number of citations a highly cited paper can get (as several of the respondents also realise). That is, when the population of papers is large the most frequently cited papers may obtain a higher number of citations compared to the similar papers in smaller fields.

[18] Other examples: "Review articles are always highly cited when published when a research field is expanding. This [highly cited] article did not contribute very much to the field, but represented (according to my colleagues) a good introduction for a post graduate student, for example. / Review articles are frequently cited because they are more easily are selected when the number of references needs to be restricted (usually 1– 2 references per assertion)."

A related issue only incidentally mentioned by the respondents is that the average citation rates vary considerably among fields. This is an issue which cannot be derived from the respondent's own scientific practice and experiences. When citation indicators are to be used in comparative assessments, normalization procedures can take the differences between fields into account. For example, when relative indicators are being constructed, the citation counts of the articles are often compared with the average field/subfield citation rate. But even within a field or subfield there are different scientific niches and specialities that may differ in publication and citation patterns, which cannot be captured easily.

Another recurrent issue concerned the effect of journals:

> An article in a highly ranked journal obtains more citations than an article of similar quality in a lower ranked journal.
> The citation counts are influenced by the accessibility of the journal.

The journal effect was also used as an explanation when the respondents commented on the citation counts of their own articles:

> This article was of little importance. But because it was published in 'Nature' it has been frequently cited.
> This article was published in an Indian journal and could not expect much attention.
> An unfortunate choice of journal. It should have been published in a more specialised journal.[19]

It is a widespread opinion that the citation count of a paper is highly influenced by the journal it is published in. In several cases the respondents seem to think that the journal is more important than the content of the article. There is a prima facie plausibility that the status of journals affects citation counts. Papers published in prestigious journals obtain more visibility and will therefore be more cited. The issue is more complex, however, because acceptance for publication in a prestigious journal is conditional on higher than average quality. This would explain the finding that the prestige of a journal is an important predictor of citations (Peters & van Raan, 1994). Others have argued that the citedness of publications is not significantly influenced by the status of the journals in which they are published, because the distribution of article citedness within most journals is very skewed, which implies that contents of the articles are the important determinants in citation rates (Seglen, 1994).

[19] Another respondent argues that: "The journal in which a paper is published significantly influences how many citations it receives. One of the most important publications I have contributed to was published in Scan J. Immunol. This article is cited but I think it is cited to a considerably lesser degree than if it had been published in a VSOP journal. In applications in which I list my most important relevant publications, I rarely mention this paper. It looks better to list articles published in Science, Nature Medicine and J. Exp. Med."

In conclusion

This study has investigated researchers' perceptions of citations. What we have as data is the repertoire of views and experiences about citations, and the paper reports and characterizes this, without too much analysis. Nevertheless, some conclusions can be drawn.

The respondents were able to mention a large number of reasons why citation counts do not correspond with scientific contribution. This was based on their own experience, as well as their referring to a shared, and somewhat cynical repertoire about citations. However, despite these problems, when asked about the citation counts of their own publications another pattern emerged. For the majority of these publications the citation counts were considered to correspond well with the degree of contribution (Aksnes, 2005). The shared repertoire may be too cynical, and lacking empirical support. It continues to be reproduced, however, because negative experiences tend to be more salient (and there is little opportunity to get a view of patterns at more aggregate levels), and there are no incentives to check the validity of the repertoire views. A questionnaire asking questions about own papers with different citation counts may introduce some reflection, as with the respondent who said:

> This investigation has made me to think through the relation between citation counts and importance. After filling in the questionnaire I am more convinced that citation counts (used sensibly) is a good indicator of importance.

Having noted this, we should add that the respondents generally appeared to be quite knowledgeable of citations. This reflects the increasing importance of citations, not only because of their use in external assessments, but also because references to citation counts (with their easy availability) are part of competitive struggles among scientists. The one limitation appears to be the lack of overview across fields. Here, bibliometrics with its link to decisions requiring comparisons across fields has developed sophisticated approaches, for example relative indicators (see e.g. Schubert, Glänzel, & Braun, 1988).

In the study we have also identified ambivalence in the resechers‘ views about citations. Citations are sought-after because they are perceived as part of the reward system of science but also criticised for not reflecting actual scientific contribution. This then leads to further ambivalences, where high citation counts are accepted as reflecting quality, and low citation counts are explained as a result of visibility and obliteration dynamics. An important finding is that this is not an across-the-board approach. In fact, overcitedness is recognized as well, and understood in terms of timeliness and relevance to larger audiences. In other words, scientists have a sophisticated understanding of the citation process and its outcomes, and can explicate such understanding when there are no immediate stakes to be defended, as in a questionnaire asking them about earlier papers.

This has implications for how to view the almost automatic resistance of scientists to measures and indicators for external assessment. One component is the scientists' wish to retain autonomy, and therefore come up with methodological and other criticisms of the indicators. (Which can actually help to improve such indicators.) From the other side (policy makers) such criticisms will then be seen as tactical ploys.

The other component relates to the experience-based expertise of the scientists, which was amply documented in this paper. Criticisms of and concerns about citation indicators should then be taken seriously, even if these have to be checked against analyst's insights (sociologists as well as bibliometricians). In this paper, we have indicated where such checks are in order.

The key lesson to be drawn out is how quality dynamics and visibility dynamics occur together in the citation process, and cannot easily be disentangled. There is no easy recipe to separate substance from tactical ploys. On the other hand, there are sophisticated insights that can and should be taken into account, by policy makers as well as by bibliometricians who want to be relevant to the complexities of the real world.

Acknowledgements

The authors are grateful to the many researchers who took the time to fill in the questionnaires.

References

Aksnes, D. W. (2003a). Characteristics of highly cited papers. *Research Evaluation, 12*(3), 159-170.

Aksnes, D. W. (2003b). A macro study of self-citation. *Scientometrics, 56*(2), 235-246.

Aksnes, D. W. (2005 (forthcoming)). Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology.*

Broad, W. & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science.* New York: Simon and Schuster.

Collins, P. M. D. (1991). *Quantitative assessment of departmental research. A survey of academics' views* (No. SEPSU Policy Study No. 5). London: Science and Engineering Policy Studies Unit of The Royal Society and The Fellowship of Engineering.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Garfield, E. (1998). From Citation Indexes to Informetrics: Is the Tail Now Wagging the Dog? *Libri, 48,* 67-80.

Garfield, E., & Welljams-Dorof, A. (1992). Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. *Science and Public Policy, 19*(5), 321-327.

Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited beyond the standard bibliometric time horizon. *Scientometrics, 58*(3), 571-586.

Luukkonen, T. (1990). *Citations in the rhetorical, reward, and communication systems of science.* Unpublished PhD thesis, University of Tampere, Tampere.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435-444.

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*, 61-90.

Martin, B. R., & Skea, J. E. F. (1992). *Academic Research Performance Indicators: An Assessment of the Possibilities*. Brighton: University of Sussex.

Merton, R. K. (1968). The Matthew Effect in Science. *Science, 159*, 56-63.

Moed, H. F., & van Leeuwen, T. N. (1995). Improving the Accuracy of Institute for Scientific Information's Journal Impact Factors. *Journal of the American Society for Information Science, 46*(6), 461-467.

Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1996). A critical analysis of the journal impact factors of 'Angewandte Chemie' and 'The Journal of the American Chemical Society'. Inaccuracies in published impact factors based on overall citations only. *Scientometrics, 37*(1), 105-116.

Mulkay, M. (1972). *The Social Process of Innovation. A Study in the Sociology of Science*. London: MacMillan.

Narin, F., Stevens, K., & Witlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics, 21*(3), 313-323.

Peters, H. P. F., & van Raan, A. F. J. (1994). On Determinants of Citation Scores: A Case Study in Chemical Engineering. *Journal of the American Society for Information Science, 45*(1), 39-49.

Price, D. d. S. (1976). A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science, 27*, 292-306.

Schubert, A., Glänzel, W., & Braun, T. (1988). Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools. In A. F. J. Van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier.

Seglen, P. O. (1994). Causal Relationship between Article Citedness and Journal Impact. *Journal of the American Society for Information Science, 45*(1), 1-11.

Seglen, P. O. (1997). Citations and journal impact factors: questionable indicators of research quality. *Allergy, 52*(11), 1050-1056.

Small, H. (2004). Why authors think their papers are highly cited. *Scientometrics, 60*(3), 305-316.

Weinberg, A. M. (1963). Criteria for Scientific Choice. *Minerva*, 1(2), 159-171.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, Forthcoming*.

Wouters, P. (1999). *The Citation Culture.* Unpublished PhD-thesis, University of Amsterdam, Amsterdam.

Samenvatting (Dutch summary)

Citaties worden in toenemende mate gebruikt als indicatoren in de context van wetenschapsbeleid en onderzoeksevaluatie. De aanname is dan dat citaties gezien kunnen worden als een maat voor wetenschappelijke kwaliteit of impact. In de afgelopen drie decades is een groot aantal studies uitgevoerd om na te gaan tot op welke hoogte deze aanname geldig is. Terwijl veel studies vonden dat citatie indicatoren redelijk goed corresponderen met de verschillende maatstaven voor onderzoeksprestatie of wetenschappelijke erkenning, waren er toch ook een aantal studies dergelijk gebruik van citaties ter discussie stelden of bekritiseerden. Met name zijn de toepassingen van citatie indicatoren voor evaluatieve doelen vaak controversieel geweest. De vraag wat gemeten wordt met behulp van citaties en de geldigheid en betrouwbaarheid van citatie indicatoren is derhalve nog steeds onderwerp van debat.

Deze disstertaie beoogt een bijdrage te leveren aan de discussie over het gebruik van citaties als indicatoren. Op basis van verschillende deelprojekten onderzoekt de dissertatie de intersectie tussen bibliometrische en sociologische vragen over het verschijnsel citaties, en over de wetenschapsbeleid kwesties over het gebruik van dergelijke data als indicatoren in besluitvorming en evaluatie. Er wordt met name ingegaan op veelgeciteerde artikelen. Citatie verdelingen zijn zeer asymmetrisch, omdat het merendeel van de artikelen nauwelijks of helemaal niet geciteerd worden, en enkele publicaties zeer vaak geciteerd worden. Het zal duidelijk zijn dat met dit verschijnsel rekening gehouden moet worden als citatie indicatoren geconstrueerd en gebruikt worden. Het is daarom interessant en belangrijk om patronen in veelgeciteerde artikelen te analyseren. Dit leidt dan vervolgens tot een aantal studies naar de methodologische basis en de geldigheid van citatie indicatoren.

De dissertatie bestaat uit een bundeling van zeven artikelen. Vijf daarvan zijn in tijdschriften gepubliceerd, één is geaccepteerd voor publicatie, en de laatste zal ingediend worden voor publicatie. Deze kern van artikelen wordt voorafgegaan door twee inleidende hoofdstukken.

Het promotie-onderzoek is uitgevoerd in het Noorse Instituut voor Studies van Onderzoek en Hoger Onderwijs (NIFU STEP) in Oslo. Voor de empirische basis van het onderzoek kon gebruik gemaakt worden de bibliometrische databestanden welke NIFU STEP gekocht heeft van Thomson ISI. Verder werden data verzameld via questionaires. Alle studies in deze dissertatie zijn gebaseerd op data voor Noorwegen.

Het openingshoofdstukgeeft een kort overzicht van de literatuur over citaties, met name vanuit een onderzoeksbeleid perspectief. Omdat citatiestudies bijna zonder uitzondering afhankelijk zijn van, en gebaseerd op data van ISI wordt in dit hoofdstuk ook de aard van het databestand van ISI beschreven. Vervolgens wordt ingegaan op de kenmerken van citatie-verdelingen en manieren om citatie-indicatoren te construeren. Tenslotte worden de kwesties besproken van de interpretatie van citatie-indicatoren en de beperkingen en problemen van citatie-indicatoren, waarbij ook voorbeelden gegeven worden van het gebruik van citaties als indicatoren in wetenschapsbeleid en onderzoeksbeoordeling.

In hoofdstuk 2 worden de centrale kwesties van asymmetrische citatie verdelingen en veelgeciteerde artikelen (en het gebruik van citaties als indicatoren) meer gericht besproken. Dit vormt dan het startpunt voor een overzicht van de hierna volgende hoofdstukken en een korte bespreking van data en onderzoeksmethodologie. De hoofdstukken 3-9 bevatten de zeven artikelen die de kern van de dissertatie vormen.

In de eerste studie (artikel I) worden de citatie-verdelingen die achter nationaal-niveau citatie-indicatoren liggen onderzocht. Het voornaamste doel is na te gaan hoe het citatie-gemiddelde beïnvloed wordt door enkele veelgeciteerde artikelen. Het bleek dat de gemiddelde mate van geciteerdheid (*citation rate*) in de voornaamste deelgebieden representatieve voor de Noorse wetenschap sterk bepaald werd door één of enkele veelgeciteerde artikelen. In veel van de gebieden waren vijf artikelen van een totale ourput van meer dan honderd artikelen verantwoordelijk voor ongeveer de helft van de citaties. De implicaties van deze bevindingen worden besproken vanuit het perspectief van onderzoeksbeleid.

In de tweede studie van de dissertatie (artikel II) werden de bibliometrische kenmerken van veelgeciteerde Noorse onderzoeksartikelen geanalyseerd. Deze blijken sterk te verschillen van "gewone" artikelen. Typerend was dat veelgeciteerde artikelen een groot aantal auteurs hadden, met vaak internationale samenwerking. De meerderheid (81%) van deze artikelen waren reguliere tijdschrift-artikelen, maar review artikelen waren met 12% oververtegenwoordigd vergeleken met het gemiddelde in de verzameling van alle Noorse publicaties.

De derde studie van de dissertatie (artikel III) beoogt bij te dragen aan onze kennis van zelf-citaties. Een zelf-citatie (een auteurs-zelf-citatie) is een citatie waarin een auteur zichzelf citeert, dat wil zeggen een eerder door hem gepubliceerd werk. Als citaties voor indicatoren worden gebruikt ziet men zelf-citaties veelal als een probleem. In de studie werden meer dan een half miljoen citaties naar de wetenschappelijke produktie van Noorwegen in de periode 1981-1996 geanalyseerd. Wanneer een drie-jaar citatie-venster gehanteerd wordt blijkt gemiddeld meer dan een derde van de citaties auteurs-zelf-citaties te zijn. Als een langere periode gebruikt werd om citaties te traceren nam dit aandeel significant af. Het hoogste voorkomen van zelf-citaties werd aangetroffen bij de minst geciteerde artikelen. Er was ook een sterke positieve correlatie tussen het aantal zelf-citaties en het aantal auteurs van een publicatie. De implicaties van deze bevindingen werden besproken.

Afbakening van gebieden en de vraag hoe dekkend data-bestanden zijn zijn de voornaamste kwesties die in de vierde studie opgepakt worden (artikel IV). De meeste bibliometrische analyses maken gebruik van op tijdschriften gebaseerde definities van deelgebieden, dat wil zeggen dat alle artikelen in een gegeven tijdschrift aan hetzelfde deelgebied worden toegekend. In de studie analyseer ik hoe goed een op tijdschriften gebaseerde onderwerp-classificatie de nationale produktie in een bepaald gebied – microbiologie – voorstelt. Het overgrote deel (94%) van de internationale wetenschappelijke produktie in microbiologie in Noorwegen werd gedekt door het bibliometrisch databestand *National Science Indicators* (NSI). Slechts enkele tijdschriften misten en deze waren waarschijnlijk niet belangrijk voor microbiologie. Daar stonc echter tegenover dat slechts 41% van de publicaties die experts classificeerden als microbiologie ook terug te vinden waren onder het etiket ‘microbiologie’ zoals door NSI gehanteerd wordt. De verzameling tijdschriften die het gebied definieerde was dus niet voldoende om dit complexe biomedische gebied te omlijnen. Dit betekent dat het niet gebrek aan dekking van het gebied is, maar tekortschieten van de afbakening van het gebied dat voor methodologische problemen zorgt in bibliometrische analyse, althans voor microbiologie.

Ter validering zijn citatie-indicatoren vergeleken met resultaten van verschillende soorten peer review. De aanname is immers dat er correlatie moet zijn als citaties inderdaad wetenschappelijke prestatie (*performance*) indiceren en dus terecht als indicatoren gebruikt mogen worden. Aan deze kwestie levert de vijfde studie (artikel V) een verdere bijdrage door de relatie te onderzoeken tussen bibliometrische indicatoren en de uitkomst van een peer review in de Universiteit van Bergen, Noorwegen. Een centrale vraag is hoe de scores die peers geven voor de wetenschappelijke prestatie van onderzoeksgroepen correleren met diverse bibliometrische indicatoren. Citatie-indicatoren staan in deze studie voorop, maar andere soorten bibliometrische indicatoren zijn ook geanalyseerd. Positieve maar vrij zwakke correlaties werden gevonden voor alle bestudeerde indicatoren. Vervolgens is nagegaan wat de redenen konden zijn voor discrepantie. Zowel tekortkomingen in de evaluatie door de peers, las in de bibliometrische indicatoren, als een gebrek aan vergelijkbaarheid speelden een rol in de verklaring waarom de overall correlatie niet sterker was.

In de zesde studie (artikel VI) analyseer ik de correspondenties tussen de citatie-tellingen van publicaties en de evaluatie die de auteurs zelf maken van hun bijdragen. De studie is gebaseerd op een survey via vragenlijsten toegestuurd aan de auteurs van de artikelen die eerder geïdentificeerd waren als veelgeciteerd. Er bleek redelijke correspondentie te zijn tussen de citatie-tellingen van de tien aan de auteurs voorgelegde artikelen (met verschillende mate van citatie) en de evaluatie van de bijdrage aan de wetenschap die de auteurs zelf maakten. In het algemeen was de correspondentie het duidelijkst, en konden citatie-tellingen dus de bijdrage identificeren, bij majeure dan wel geringe bijdragen. Uit de auteurs oordelen blijkt wel dat citaties geen betrouwbare indicator van bijdrage aan de wetenschap zijn op het niveau van individuele artikelen. In de studie werd ook bevestigd dat review artikelen vaker geciteerd worden dan andere typen publicaties. Vergeleken met de betekenis die de auteurs zelf er aan toekennen kan dit als aanmerkelijke over-citatie gezien worden.

De laatste studie (artikel VII) onderzoekt in detail hoe citaties en citatie-tellingen gezien worden door wetenschapsmensen, gebaseerd op de questionaire survey waar artikel VI ook op gebaseerd was. De respondenten, auteurs van tenminste één veelgeciteerd artikel, waren gevraagd commentaar te geven op kwesties als waarom, of waarom niet, de citatie-telling de wetenschappelijke bijdrage van een artikel weerspiegelt (in het algemeen, en voor elk van de voorgelegde eigen artikelen), en hoe artikelen veelgeciteerd raken. Respondenten zijn ambivalent over het gebruik van citatie-indicatoren, en signaleren divergenties tussen kwaliteitsdynamieken en zichtbaarheidsdynamieken in hoe artikelen citaties verwerven.

References

Abt, H. A. (2000). Do important papers produce high citation counts? *Scientometrics, 48*(1), 65-70.

Amsterdamska, O., & Leydesdorff, L. (1989). Citations: Indicators of Significance? *Scientometrics, 15*(5-6), 449-471.

Bonzi, S., & Snyder, H. W. (1990). Patterns of self citation across fields of inquiry. *Proceedings of the ASIS annual meeting, 27*, 204-207.

Bourke, P., & Butler, L. (1996). Publication types, citation rates and evaluation. *Scientometrics, 37*(3), 437-494.

Braun, T., Glänzel, W., & Schubert, A. (1989). Assessing assessments of British science - some facts and figures to accept or decline. *Scientometrics, 15*(3-4), 165-170.

Braun, T., Glänzel, W., & Grupp, H. (1995). The scientometric weight of 50 nations in 27 science areas, 1989-1993. Part 1. All fields combined, mathematics, engineering, chemistry and physics. *Scientometrics, 33*(3), 263-293.

Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation, 12*(1), 39-46.

Callon, M., Law, J., & Rip, A. (Eds.). (1986). *Mapping the Dynamics of Science and Technology*. London: Macmillan.

Campanario, J. M. (1993). Consolation for the Scientist: Sometimes it is Hard to Publish Papers that are Later Highly-Cited. *Social Studies of Science, 23*, 342-362.

Campanario, J. M. (1996). Using Citation classics to study the incidence of serendipity in scientific discovery. *Scientometrics, 37*(1), 3-24.

Cano, V., & Lind, N. C. (1994). Citation Life Cycles of Ten Citation Classics. *Scientometrics, 22*(2), 297-312.

Case, D. O., & Higgins, G. M. (2000). How Can We Investigate Citation Behavior? A Study of Reasons for Citing Literature in Communication. *Journal of the American Society for Information Science, 51*(7), 635-645.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless Science. Peer Review and U.S. Science Policy*. Albany: State University of New York Press.

Cole, F. J., & Eales, N. B. (1917). The history of comparative anatomy. *Science Progress, XI*, 578-596.

Cole, J. R., & Cole, S. (1972). The Ortega Hypothesis. *Science, 178*, 368-375.

Cole, J. R., & Cole, S. (1973). *Social Stratification in Science*. Chicago: The University of Chicago Press.

Cole, S. (1970). Professional Standing and the Reception of Scientific Discoveries. *American Journal of Sociology, 76*, 286-306.

Cole, S. (1992). *Making Science. Between Nature and Society*. London: Harvard University Press.

Cole, S. (2000). The Role of Journals in the Growth of Scientific Knowledge. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Medford: ASIS.

Collins, R. (1975) *Conflict Sociology. Toward an Explanatory Science.* New York: Academic Press.

Cozzens, S. E. (1989). What do citations count? The rhetoric-first model. *Scientometrics, 15*(5-6), 437-447.

Cozzens, S. E. (1997). The Discovery of Growth: Statistical Glimpses of Twentieth-Century Science. In J. Krige & D. Pestre (Eds.), *Science in the Twentieth Century* (pp. 127-142). Amsterdam: Harwood Academic Publishers.

Cronin, B. (1984). *The Citation Process. The Role and Significance of Citations in Scientific Communication*. London: Taylor Graham.

Davenport, E., & Cronin, B. (2000). The Citation Network as a Prototype for Representing Trust In Virtual Environments. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Medford: ASIS.

Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science.* Amsterdam: Elsevier Science Publishers.

Elkana, Y., Lederberg J., Merton, R. K., Thackray A., Zuckerman, H. (Eds.) (1978). *Toward a Metric of Science. The Advent of Science Indicators.* New York: John Wiley & Sons.

European Commission. (2001). *Key Figures 2001. Special edition. Indicators for benchmarking of national research policies*. Brussels.

Garfield, E. (1977). Can Citation Indexing Be Automated? In *Essay of an Information Scientist, vol. 1* (Vol. 1). Philadelphia: ISI Press.

Garfield, E. (1979a). *Citation Indexing - Its Theory and Application in Science, Technology and Humanities*. New York: John Wiley & Sons.

Garfield, E. (1979b). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Garfield, E. (1980). The Number of Biochemical Articles Is Growing, But Why Also the Number of References per Article. *Current Contents*(11), 5-9.

Garfield, E. (1990). The most-cited papers of all times, SCI 1945-1988. The SCI top 100 - Will the Lowry method ever be obliterated. *Current Contents*(7), 3-15.

Garfield, E. (1997a). Dispelling a Few Common Myths about Journal Citation Impacts. *The Scientist*, *11*(33), 11.

Garfield, E. (1997b). Validation of Citation Analysis. *Journal of the American Society for Information Science*, *48*(10), 962-963.

Garfield, E. (1998a). From Citation Indexes to Informetrics: Is the Tail Now Wagging the Dog? *Libri*, *48*, 67-80.

Garfield, E. (1998b). *The use of journal impact factors and citation analysis for evaluation of science*. URL: http://www.garfield.library.upenn.edu/papers/eva_of_science_oslo.html

Garfield, E., & Welljams-Dorof, A. (1992). Of nobel class: A citation perspective on high impact research authors. *Theoretical Medicine*, *13*(2), 117-135.

Gilbert, N. G. (1977). Referencing as Persuasion. *Social Studies of Science*, *7*, 113-122.

Glänzel, W., & Czerwon, H.-J. (1992). What are highly cited publications? A method applied to German scientific papers, 1980-1989. *Research Evaluation*, *2*(3), 135-141.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171-193.

Glänzel, W., Rinia, E. J., & Brocken, M. G. M. (1995). A bibliometric study of highly cited European physics papers in the 80s. *Research Evaluation*, *5*(2), 113-122.

Glänzel, W., & Schubert, A. (1992). Some Facts and Figures on Highly Cited Papers in the Sciences, 1981-1985. *Scientometrics*, *25*(3), 373-380.

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science, LXVI*(1713), 385-389.

Hamilton, D. P. (1990). Publishing by - and for? - the Numbers. *Science*, *250*(4986), 1331-1332.

Hamilton, D. P. (1991). Research Papers: Who's Uncited Now? *Science*, *251*(4989), 25.

Hargens, L. L. (2000). Graphing Micro-Regions in the Web of Knowledge: A Comparative Reference-Network Analysis. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield* (pp. 497-516). Medford: ASIS.

Hargens, L. L., & Schuman, H. (1990). Citation counts and social comparisons: Scientists' use and evaluation of citation index data. *Social Science Research*, *19*(3), 205-221.

Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, *44*(2), 193-215.

Irvine, J., Martin, B., Peacock, T., & Turner, R. (1985). Charting the Decline in British Science. *Nature*, *316*(6029), 587-590.

Katz, S. J., & Hicks, D. (1998). *Indicators for Systems of Innovation* (No. IDEA 12-98). Oslo: STEP-group.

Kyvik, S. (1991). *Productiviy in Academia. Scientific publishing at Norwegian universities*. Oslo: Universitetsforlaget.

Law, J. (1986). The Heterogeneity of Texts. In M. Callon, J. Law & A. Rip (Eds.), *Mapping the Dynamics of Science and Technology*. London: Macmillan.

Lawani, S. M. (1982). On the Heterogeneity and Classification of Author Self-Citations. *Journal of the American Society for Information Science*, *33*, 281-284.

Leeuwen van, T. N., Rinia, E. J., & van Raan, A. F. J. (1996). *Bibliometric Profiles of Academic Physics Research in the Netherlands*. Leiden: Centre for Science and Technology Studies. Report CWTS 96-09.

lePair, C. (1995). Formal evaluation methods: Their utility and limitations. *International Forum on Information and Documentation, 20*(4), 16-24.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317-323.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randal, R. J. (1951). Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry, 193*, 265-275.

Luukkonen, T. (1990). *Citations in the rhetorical, reward, and communication systems of science.* Unpublished PhD thesis, University of Tampere, Tampere.

Luukkonen, T. (1991). Citation indicators and peer review: their time-scales, criteria of evaluation, and biases. *Research Evaluation, 1*(1), 21-30.

Luukkonen, T. (1997a). Quantitative Techniques in Evaluation in Western Europe. In M. S. Fankel & J. Cave (Eds.), *Evaluating Science and Scientists. An East-West Dialoge on Research Evaluation in Post-Communist Europe.* Budapest: Central European University Press.

Luukkonen, T. (1997b). Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics, 38*(1), 27-37.

MacRoberts, M. (1997). Rejoinder: Validation of Citation Analysis. *Journal of the American Society for Information Science, 48*(10), 963.

MacRoberts, M. H., & MacRoberts, B. R. (1987). Testing the Ortega Hypothesis: Facts and Artifacts. *Scientometrics, 12*(5-6), 293-296.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of Citation Analysis: A Critical Review. *Journal of the American Society for Information Science, 40*(5), 342-349.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435-444.

Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, *36*(3), 343-362.

Martin, B. R. (1997). Factors Affecting the Acceptance of Evaluation Results. In M. S. Fankel & J. Cave (Eds.), *Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe*. Budapest: Central European University Press.

Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*, 61-90.

Martin, B. R., & Skea, J. E. F. (1992). *Academic Research Performance Indicators: An Assessment of the Possibilities*. Brighton: University of Sussex.

May, K. O. (1968). Growth and Quality of the Mathematical Literature. *ISIS, 59*,363-371

May, R. (1997). The scientific wealth of nations. *Science, 275*(5301), 793-796.

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*, 56-63.

Merton, R. K. (1979). Foreword. In E. Garfield (Ed.), *Citation Indexing - Its Theory and Application in Science, Technology, and Humanities*. New York: John Wiley & Sons.

Merton, R. K. (1988). The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property. *ISIS, 79*, 606-623.

van der Meulen, B. J. R. (1997). The use of S&T indicators in science policy. Dutch experiences and theoretical perspectives from policy analysis. *Scientometrics, 38*(1), 87-101.

Miquel, J. F., Ojasoo, T., Okubo, Y., Paul, A., & Dore, J. C. (1995). World science in 18 disciplinary areas: Comparative evaluation of the publication patterns of 48 countries over the period 1981-1992. *Scientometrics, 33*(2), 149-167.

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

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. J. (1985). The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics, 8*(3-4), 177-203.

Moed, H. F., Burger, W. J. M., Frankfurt, J. G., & van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*, 131-149.

Moed, H. F., & van der Velde, J. G. M. (1993). *Bibliometric Profiles of Academic Chemistry Research in the Netherlands*. Leiden: Centre for Science and Technology Studies. Report CWTS-93-08.
Moed, H. F., Van Leeuwen, T. N., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics*, *46*(3), 575-589.
Moed, H. F., & Vriens, M. (1989). Possible inaccuracies occuring in citation analysis. *Journal of Information Science*, *15*, 95-107.
Narin, F., Hamilton, K. S., & Olivastro, D. (2000). The Development of Science Indicators in the United States. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Meford, New Jersey: Information Today, Inc.
Nederhof, A. J., Zwaan, R. A., De Bruin, R. E., & Dekker, P. J. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social and behavioural sciences. A comparative study. *Scientometrics, 32*(1), 37-48.
Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation, 53*(5), 477-487.
Oppenheim, C., & Renn, S. P. (1978). Highly Cited Old Papers and the Reasons Why They Continue to be Cited. *Journal of the American Society for Information Science, 29*(4), 225-231.
Persson, O. (2000). *Decline of Swedish Science? (Bibliometric Notes, No3 2000).* URL: http://www.umu.se/inforsk/BibliometricNotes/BN3-2000/BN3-2000.htm.
Phelan, T. J. (1999). A compendium of issues for citation analysis. *Scientometrics, 45*(1), 117-136.
Plomp, R. (1994). The Highly Cited Papers of Professors as an Indicator of a Research Group's Scientific Performance. *Scientometrics, 29*(3), 377-393.
Price, D. J. d. S. (1965). Networks of Scientific Papers. *Science, 149*, 510-515.
Price, D. J. d. S. (1970). Citation Measures of Hard Science, Soft Science, Technology, and Nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication Among Scientists and Engineers* (pp. 3-22). Lexington: D. C. Heat & Co.
van Raan, A. F. J. (1993). Advanced bibliometric methods to assess reseach performance and scientific development: basic principles and recent practical applications. *Research Evaluation, 3*(3), 151-166.
van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and forsight exercises. *Scientometrics, 36*(3), 397-420.
van Raan, A. F. J. (2000). The Pandora's Box of Citation Analysis: Measuring Scientific Excellence -- The Last Evil? In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield* (pp. 301-319). Medford: ASIS.
van Raan, A. F. J. (2005). Fatal attraction: Ranking of universities by bibliometric methods. *Scientometrics* 62 (1), 133-143.
Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*, 95-107.
Rip, A. (1997). Qualitative conditions of scientometrics: The new challenges. *Scientometrics, 38*(1), 7-26.
Schoepflin, U. (1992). *Problems of Representativity in the Social Sciences Citation Index.* Paper presented at the Representation of Science and Technology. Proceedings of the International Conference on Science and Technology Indicators, Bielefeld, Germany, 10-12 June 1990.
Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5-6), 281-291.
Schubert, A., & Braun, T. (1993). Reference standards for citation based assessments. *Scientometrics, 26*(1), 21-35.
Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. *Scientometrics, 36*(3), 311-324.

Schubert, A., Glänzel, W., & Braun, T. (1988). Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools. In A. F. J. Van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier.

Schwartz, C. A. (1997). The Rise and Fall of Uncitedness. *College & Research Libraries, 58*(1), 19-29.

Seglen, P. O. (1989). From bad to worse: Evaluation by Journal Impact. *Trends in Biochemical Sciences, 14*, 326-327.

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

Seglen, P. O. (1994). Causal Relationship between Article Citedness and Journal Impact. *Journal of the American Society for Information Science, 45*(1), 1-11.

Seglen, P. O. (1997). Citations and journal impact factors: questionable indicators of research quality. *Allergy, 52*(11), 1050-1056.

Shadish, W. R., Tolliver, D., Gray, M., & Gupta, S. K. S. (1995). Author Judgements about Works They Cite: Three Studies from Psychology Journals. *Social Studies of Science, 25*, 477-498.

Sivertsen, G. (1992). *Should a new bibliometric database for international comparisons be more restricted in journal coverage.* Paper presented at the Science and technology in a policy context. Leiden Conference on Science & Technology Indicators,, Leiden, The Netherlands, 23- 25 October 1991.

Sivertsen, G. (1997). Ethical and Political Aspects of Using and Interpreting Quantitative Indicators. In M. S. Frankel & J. Cave (Eds.), *Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe*. Budapest: Central European University Press.

Small, H. (1982). Citation Context Analysis. In B. Dervin & M.-. Voigt (Eds.), *Progress in communication sciences* (Vol. 3, pp. 287-310). Norwood: Ablex.

Small, H. G. (1978). Cited Documents as Concept Symbols. *Social Studies of Science, 8*, 327-340.

Smith, L. C. (1981). Citation Analysis. *Library Trends, 30*(1), 83-106.

Snyder, H., & Bonzi, S. (1998). Patterns of self-citation across disciplines (1980-1989). *Journal of Information Science, 24*(6), 432-435.

Tagliacozzo, R. (1977). Self-citation in scientific literature. *Journal of Documentation, 33*(4), 251-265.

Taxt, R. E., & Aksnes, D. W. (2003). *Vitenskapelig publisering i perioden 1995-2000. En analyse av Det matematisk-naturvitenskapelige fakultetet*. Bergen: Universitetet i Bergen.

Testa, J. (1997). The ISI Database: the journal selection process. *The ISI Essays. URL: http://www.isinet.com/isi/hot/essays/199701.html*.

Tijssen, R. J. W. (2003). Scoreboards of research excellence. *Research Evaluation, 12*(2), 91-103.

Tijssen, R. J. W., Visser, M. S., & van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381-397.

Turner, S. P., & Chubin, D. E. (1976). Another appraisal of Ortega, the Coles, and science policy: the Ecclesiastes hypothesis. *Social Science Information, 15*(4/5), 657-662.

Turner, S. P., & Chubin, D. E. (1979). Chance and eminence in science: Ecclesiastes II. *Social Science Information, 18*(3), 437-449.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics, 10*(3-4), 157-177.

Vinkler, P. (1998). Comparative investigation of frequency and strength of motives toward referencing. The referencing threshold model. *Scientometrics, 43*(1), 107-127.

Voos, H., & Dagaev, K. S. (1976). Are All Citations Equal? Or, Did We Op. Cit. Your Idem? *The Journal of Academic Librarianship, 1*(6), 19-21.

VSNU. (1996). *Quality Assessment of Research - Chemistry, past performance and future perspectives*. Utrecht: VSNU.

Warner, J. (2000). A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science, 26*(6), 453-460.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, Forthcoming*.
Welljams-Dorof, A. (1997). Quantitative Citation Data as Indicators in Science Evaluations: A Primer on Their Appropriate Use. In M. S. Frankel & J. Cave (Eds.), *Evaluating Science and Scientists. An East-West Dialogue on Research Evaluation in Post-Communist Europe.* Budapest: Central European University Press.
Wouters, P. (1999). *The Citation Culture.* Unpublished PhD-thesis, University of Amsterdam, Amsterdam.
Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics, 12*(5-6), 329-338.