Dag W. Aksnes

Output and Effects of R&D

Science and Technology Output Indicators: An overview

NIFU skriftserie nr. 19/98

NIFU - Norsk institutt for studier
av forskning og utdanning
Hegdehaugsveien 31
0352 Oslo

ISSN 0808-4572

Preface

This report is one result of the initial, preparatory studies which are being done at NIFU within the framework of a new strategic research programme on "Profiling Output in Norwegian Research" (1998 - 2001). The programme is initiated to enhance information and knowledge on the output and effects of Norwegian research. The programme is financed by the Research Council of Norway. Together with a report to be published shortly on output profiles in the Norwegian research institute sector, this report provides part of a state-of-the-art assessment of the area of R&D output indicators, discussing i.a. the opportunities, challenges and problems raised by these kinds of output indicators and analyses.

Oslo, December 1998

Petter Aasen
Director

Egil Kallerud
Research Director

Contents

Introduction

Traditionally the major area for indicators in science policy has been related to the input side, for example indicators on R&D expenditures and R&D personnel. The last two decades, however, there has been an increasing focus on developing indicators for measuring the output of science and technology. This report attempts to give a general overview of the situation concerning science and technology (S&T) output indicators, both in an international and national context.

The study of the results and effects of R&D may involve both quantitative and qualitative approaches. As the title suggests, this report will mainly concentrate on quantitative studies on the output of R&D. Furthermore, the main focus lies on indicators related to the macro level, that is, measurements and analyses of the results of R&D on a national level. Both science and technology indicators are included. The main focus will, however, be on science indicators related to the results of basic and applied research, not on indicators of technological output.

The first chapter of the report gives a general overview of the different types of S&T output indicators. In Chapter 2 some examples of the use of S&T output indicators in science policy are further described. In particular, this chapter discusses the possibilities and the experiences of such kinds of applications of indicators. Chapter 3 contains a short description of previous studies in the area of S&T output indicators in Norway.

1 Science and technology output indicators – a general overview

1.1 Introduction

Generally speaking, science and technology indicators are developed on the basis of quantitative data on different aspects of the research and innovation system. As indicators they are normally used for the purpose of making comparisons, for example between countries or sectors and across several years. Thus a typical example will be how the national research performance is compared to other countries. Information and knowledge of the S&T system are necessary for making strategic decisions in science policy, and in this respect indicators are one of the main sources of information.

The OECD has had a key role in developing S&T indicators,[1] particular on R&D resources. An important result of this work has been an international standard for measuring R&D resources, the so-called *Frascati Manual* (OECD 1994a, first published in 1963). Since the 1960s an increasing number of techniques for measuring different aspects of the science and technology system have been developed. While the *Frascati Manual* mainly has been, and still is, devoted to input indicators, for example indicators on expenses on R&D and R&D personnel, a growing interest has been directed towards indicators on the results of R&D, in particular during the last 10-15 years. One important field has been the study of scientific publishing and the development of bibliometric indicators. Another important source of information has been surveys on innovation activities. A standard for innovation surveys was first published in 1992, known as the "Oslo Manual" (OECD 1997a).

In addition to the OECD, the National Science Foundation (NSF) in the USA has been an important organisation regarding the development of S&T indicators. In 1972 the first report on science indicators was published by the National Science Board (Cozzens 1997). The basic idea behind the report was to give an analytical quantitative description of the US science and engineering system. The report has continuously been further developed, and now contains indicators on a wide range of areas, including education, R&D resources, collaboration, technology and the public understanding of science (National Science Board 1996; National Science Board 1998). Similar reports are now issued in several countries, including Norway (Norges forskningsråd 1997).

[1] In the report the terms "R&D indicators" and "S&T indicators" are used more or less synonymously. However, it should be recognised that the term S&T indicators normally would have a wider range than R&D indicators (because technology and innovations are not necessarily a product of R&D).

Great Britain, the Netherlands, and Germany are other examples of countries that have played a prominent role in the development of S&T output indicators. To mention a few organisations, the Centre for Science and Technology Studies (CWTS) in the Netherlands has had a leading role in the development of bibliometric indicators. SPRU at the University of Sussex, PREST at the University of Manchester, MERIT at Maastricht University, and the Fraunhofer Institute in Karlsruhe, are other prominent groups in the area of S&T indicators (for a further review see e.g. Barré 1997). The last 10 years S&T studies have also been developed by the European Commission, and two indicator reports have so far been issued (European Commission 1994, European Commission 1997).

The increasing interest in science and technology indicators is related to several changes in science policy the last decades. One important factor since the 1960s has been the general demand for information and knowledge which could give policy decisions related to science a rational basis (see e.g. Edge 1995:7).

More generally, there has been a requirement from governments for greater public accountability in all areas of public expenditure, including science. With this come demands for evaluation and for performance indicators to assure governments that public money is being well spent.

In addition, stronger strategic planning of scientific activities has been necessary in many countries. Many industrialised countries are witnessing increasing constraints on public expenditure, also on expenditures on research. In consequence, it is becoming more difficult to raise the funds needed to support new areas and new scientists or pay for more sophisticated instrumentation. With an essentially level budget, reductions in existing commitments have to be made if support for new areas and researchers is to be found. The traditional peer review system has in some cases turned out to be less satisfactory when it comes to identifying declining areas and groups (Martin 1996). Indicators may thus yield valuable information in a situation which calls for strategic policy decisions.

Generally, the methodology and concepts for measuring outputs of R&D have not been standardised in a corresponding way to the more traditional input indicators. A main challenge is that it is quite difficult to measure and to quantify the results and effects of R&D in a meaningful way. Still much work has been done and a variety of indicators have been developed, involving different kinds of data, concepts and methods. Broadly speaking in the case of academic science, this in particular has involved different kind of bibliometric data focusing on scientific publications and citations. In the area of technology, patents and innovation surveys have had a key role. In addition, different kinds of indicators and analyses of the economic effects of science and technology have been developed.

The output of R&D will vary between different sectors of the R&D system. Furthermore, output and effects may be related to different kinds of "arenas". In the "output research compass" (Figure 1), a schematic overview of different kinds of arenas and possible outputs is presented (adapted by A. Kaloudis from Laredo et al. 1992).

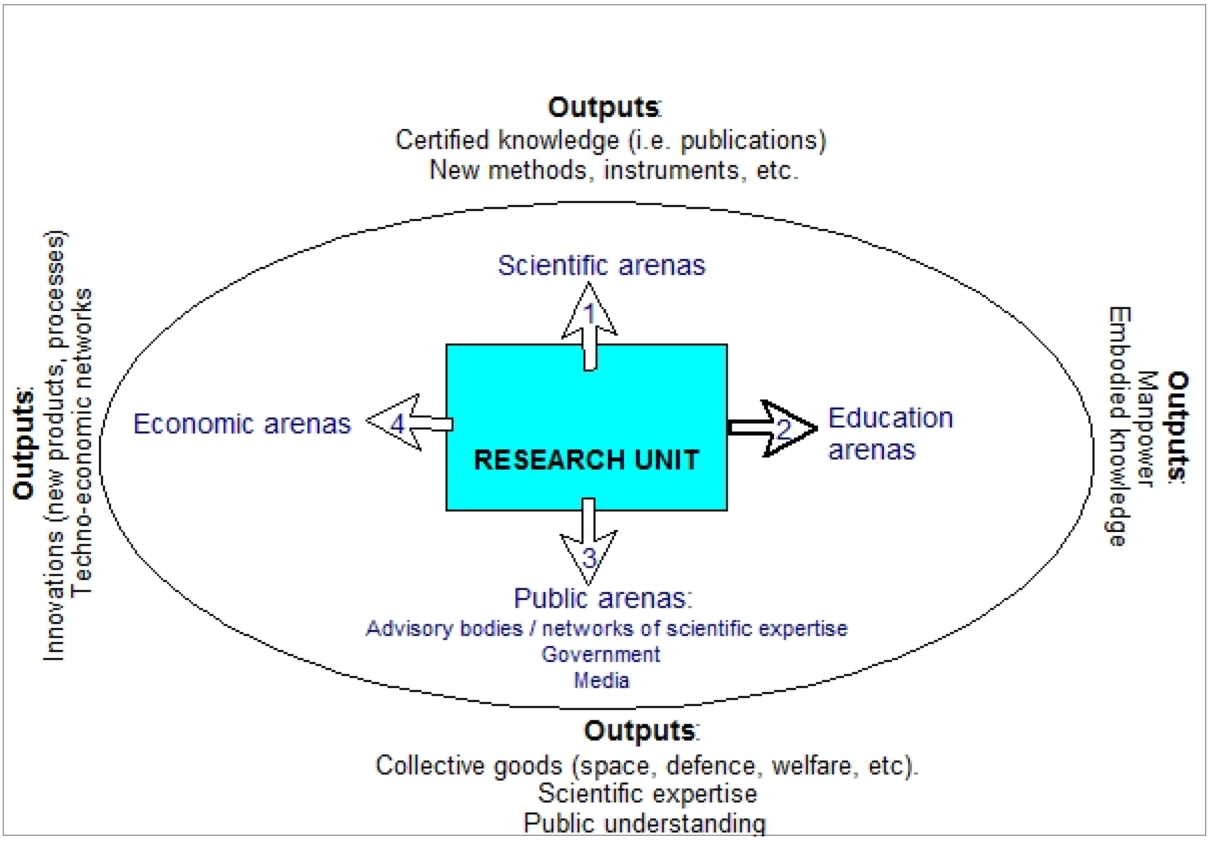

Figure 1 The output "research compass"

The study of S&T indicators can generally be said to be based on a presupposition that the results of science can be measured in a few tangible products, such as publications and patents. Such an assumption is, of course, not without problems. By treating the results of science in a quantitative manner, important aspects related for example to the sociology and conceptual progress of science are ignored. Before describing R&D output indicators in further details, it is therefore necessary to stress the obvious limitations that are related to such studies.

In the next sections the main types of indicators are briefly described. A distinction is made been indicators of the results of R&D, of scientific impact, and of the effects of R&D. Although these distinctions may sometimes be difficult to apply in practice, they nevertheless indicate important differences between sets of indicators.

1.2 Results of R&D – main indicators

1.2.1 Publication output

New knowledge is the central aim of all basic and applied research (e.g. facts, theories, and methods). However, knowledge is not an entity which is easily measured. To get information on the production of knowledge, it is therefore necessary to employ indirect indications of such production. The main indicator in this respect has been publications. The core assumption of using publications as the unity of measurement for knowledge production is that scientists tend to publish their results. The number of publications can thus be regarded as an indirect measurement of the production of research, within a country, scientific field or institution.

Scientific papers are registered and catalogued in many different databases which can be used for bibliometric purposes. The most important database, which is continuously updated, is that compiled by the Institute for Scientific Information (ISI) in Philadelphia, Pennsylvania in the United States. This database covers 16,000 specialised and multidisciplinary journals in science, medicine and technology, in addition to the social sciences and humanities (Garfield 1997). On the basis of this database different bibliometric products are produced. The most famous is the Science Citation Index (SCI), covering the natural sciences, medicine and technology. Similar bibliometric products for the social sciences and the humanities are the Social Sciences Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI). Among the conditions for extending ISI coverage to include a particular journal are refereeing, international editorial conventions and geographic representation, as well as the availability of an English language abstract. In general, the ISI database is regarded as constituting a satisfactory representation of international mainstream research (Katz & Hicks 1998).

However, researchers produce many other types of publications than articles in international scientific journals (books, reports and newspaper articles, etc.). Quantitative analyses of this kind of literature are, however, difficult because there are no databases which include these types of publications on a world-wide level. Thus bibliometric indicators are most appropriate when publications in international journals are the main carriers of knowledge. In the social sciences and humanities a large part of the production consists of books and articles in local journals. Because of this the use of bibliometric indicators in these areas is generally problematic. (This factor will, however, vary between fields and may be less problematic in “science-like” fields such as economics and psychology).

Bibliometric records of research publications in a database may include the title of the publications, citation data, the names of co-authors as well as institutional addresses. This information may be used for different kinds of analyses and international comparisons. In addition to scientific impact (see section 1.3), main areas of investigation include:

- Scientific output of researchers, institutes, universities and countries
- Development within specific disciplines world-wide or in a country
- Communication structures and collaboration (co-authorship)

The indicators can be used for monitoring purposes, in addition to assessing the strengths and weaknesses of the national research system. As indicated, bibliometric analyses may be related to different levels in the research system. Typical examples would be analyses of publication output within departments, faculties or universities (see e.g. Irvine & Martin 1984; Carpenter et al. 1988; Nederhof, Leeuwen, & Visser 1997), or within disciplines at a national level (see e.g. Moed & Velde 1993; Leeuwen, Rinia, & Van Raan 1996) or between disciplines and nations world-wide (see e.g. Braun, Glänzel, & Grupp 1995; Miquel et al. 1995)

Number of publications
The most basic bibliometric output indicator is based on counting the number of publications produced by the research unit under study. Such indicators may be expressed in absolute numbers or in relative terms. In this way the research dynamic of a given country, discipline, or institution can be monitored and its trend tracked over time. For example, the share of world publication output within a particular scientific field may be readily determined. The share of a given nation in world publications is generally regarded as a useful indicator of the productivity of its researchers. Such pictures may be of great importance for the monitoring and assessment of research. However, analyses on lower levels of aggregation are necessary to detect specific changes in scientific and technological performance.

Relative specialisation
Based on publication counts within different scientific fields, it is possible to develop indicators on the specialisation of research within a country. Such indicators are constructed by dividing the country's share of the world's publication output in a given field with the country's total share of the world's publication output (see Schubert, Glänzel, & Braun 1988). This tells us if a country has a greater percentage of its scientific paper production in this particular field than its average share within all scientific fields. In this way one gets an indication of which fields a particular country has relative specialisation in (such indicators are for example included in the European Commission 1997).

Productivity

The large differences between countries in the levels of scientific activities mean that it is often necessary to adjust for differences in input variables. Such indicators of *productivity* may be obtained by dividing the number of publications with the number of researchers (or alternatively R&D person-years) or amount invested (for an example of this see the Netherlands Observatory of Science and Technology 1996). However, the number of researchers or amount invested in a particular field within a country is neither a standard indicator nor a trivial task to calculate. Alternatively, productivity can be calculated as the number of publications per million inhabitants (see e.g. Norges forskningsråd 1997). A problem with population normalisation is, however, that the size of populations does not necessarily correspond to national differences in R&D input variables. As a rough measurement of differences in productivity such indicators may, nevertheless, give some information.

Relational bibliometric indicators

Particular relational indicators have been designed for the study of interactions within the research system. Strictly speaking they are not indicators of the output of research and will thus only be briefly mentioned here.

A major area has been the study of co-authorship (see e.g. Melin 1997). In such studies co-authorship of papers is used as a measure of scientific co-operation. For example, through analysis of co-authorship it is possible to identify collaboration at a national or international level (e.g. internationalisation).

Relational indicators are also developed on the basis of co-citation and co-word analyses. Such indicators can for example identify the network of documents within a specific field or research topic (see e.g Callon, Law, & Rip 1986; Leydersdorff 1995). Co-citation analyses use information about citation patterns to identify such structures. Co-word analyses make use of content related to bibliometric information (such as key words, words in the title or in the abstract of the selected set of publications).

Limitations with bibliometric output indicators

There are, however, many limitations with bibliometric output indicators. Although large in terms of volume and scope, the ISI databases are not necessarily a good reflection of scientific publication activity. The databases only cover a certain amount of the journals issued. This means that the number of publications registered in these databases is incomplete, and, hence, an inaccurate measurement of the actual amount of scientific production. As we have seen, publications in local journals, books, articles in newspapers and other popular literature, etc., are also not recorded.

Because of this one cannot use these databases as a measurement of the total scientific production. That is, an analysis based on international databases will only give a partial picture of the total scientific productivity of an institution or country. The importance of these factors, however, will vary from field to field and from country to country.[2] For example, it has been showed that important European journals are not included in the ISI database and that it contains more minor US journals than minor European journals. Furthermore, it is known to be strongly skewed in favour of English-language journals and research publications from English-speaking nations dominate the database (particularly the US).

The selection of journals is, nevertheless, based on an extensive evaluation process. This process is meant to ensure that the journals included have a certain international standard or impact. But as indicated above, it is questionable if one is actually comparing a representative selection of the publications from different countries.

Using other databases than the one provided by ISI, it may be possible to get a more complete picture of the total scientific production, for example local databases within a country or at university level (e.g. FORSKDOK at the University of Bergen), or specialised fields of specific bibliometric databases like MEDLINE and Chemical Abstracts. The first kind of databases may give an overview of the total scientific production, not only in refereed international journals. However such databases are normally not adjusted for bibliometric investigations and do not allow international comparisons. The latter databases may be used for field-specific bibliometric investigations, but these do not allow inter-field comparisons. Furthermore they usually lack bibliometric records like citations and complete fields of addresses.

Variable degree of coverage between different kinds of research is another bias of the ISI database. There is a rather strong focus on fundamental research, especially in the natural and life sciences. The database has been criticised for not covering technical and applied fields very well.

Subject classification is another problematic issue in bibliometric data processing. Since comparing scientific subfields on a national or international level is a major area of bibliometrics, it is important that the classification used has a valid foundation. The common procedure of delimiting scientific subfields is based upon a classification of scientific journals ('journal categories'). This means that all the

[2] For example the degree of coverage in the SCI varies among fields. It has been estimated that in chemistry about 90 per cent of the relevant journals are included in the database, compared to 30 per cent for biology (Moed et al. 1987).

articles in a given journal are assigned to one (or more) particular subfield. Most of the traditional bibliometric analyses involving subfield comparisons are based upon this kind of classification method. The method gives an indication of the productivity within different scientific fields, but is not without problems (see e.g. Aksnes, Olsen, & Seglen 1998). Generally the problems with subfield classifications are recognised as more problematic in smaller than larger subfields (Bruin & Moed 1993).

In addition to the problems mentioned above, several other methodological problems exist. Generally, because of these limitations, the use of bibliometric indicators is normally regarded as more suitable for macro and international comparisons. Furthermore, bibliometric analyses are most successful in internationally oriented fields like medicine and the natural sciences.

Not all research leads to publications. Researchers working in companies will sometimes not publish their result in public journals because of commercial interests related to the research. Because of the multi-faceted nature of research, publication indicators will never reveal more than part of the picture. Additional output indicators are thus necessary to get a more complete picture.

1.2.2 Technology and innovation indicators

While new knowledge may be regarded as the main target of basic and applied research, new or improved products and processes are important results of developmental work. According to the *Frascati Manual*, experimental development is systematic work directed to producing new materials, products or devices, to installing new processes, systems and services, or to improving substantially those already produced or installed. Most of this kind of R&D is performed in the business enterprise sector.

Technology and its specification in the form of innovation have emerged as the most important topics of economic development. While in earlier decades comparative advantages were based on natural resources and/or cheap labour costs, technology becomes the cornerstone for building competitive advantages and positioning countries in the global economy. Indicators on the extent and importance of innovations have, therefore, been an area of major interest. Among others, an important source of information has been special innovation surveys in the business enterprise sector.

Innovation analyses are one way to explore the effects of new knowledge. Innovation processes are complicated. Many different factors may be of importance. New knowledge produced through R&D may be applied in the innovation process, but is not a necessary precondition for innovation. The focus on innovations is related to a shift in the understanding of the relationship between science and technology. In the past it has often been assumed that scientific discoveries are a precondition for innovation. However, this relationship has turned out to be far more complex. The creation of innovation does not result simply from a transfer of knowledge from the science system into applications. Such insights have been an important factor in changing the focus and approaches in science and technology policy.

An OECD standard for surveys of innovation activity was first published in 1992 and is known as the "Oslo Manual" (OECD 1997a). This manual attempted to present a framework for the development of new innovation indicators which could provide a basis for international comparability.

This OECD methodology was subsequently developed by EUROSTAT and DG XIII (European Innovations Monitoring System) within the European Commission, and implemented on a European basis using a common questionnaire; this survey was known as the CIS action. EUROSTAT built a comprehensive firm-level database with the CIS data which contains data on almost 41,000 European firms. In 1997 the second CIS exercise was initiated (Sirilli 1998).

1.2.3 Patents

A particular field of study has emerged on the basis of patent records provided by patent offices. Firms, and other institutions engaged in innovative activity apply for patents to secure proprietary rights for the use of innovations. In consequence patenting data will represent an indication of the extent of technological activity results in usable outputs, products and processes.

Patent statistics have increasingly been used in various ways as indicators of the output of innovation activities. Data on the patenting activities may be provided by different offices, for example the US Patent Office which registers the number of awarded patents in the USA and the European Patent Office which registers patent applications. Different kinds of patents are registered: domestic patents (patents registered in each country by domestic inventors), and external patents (patents registered in countries other than the country of the inventor). On the basis of such data it is possible to construct indicators that, among other things, can be used as an assessment of technological specialisation vis-á-vis other nations and give hints about technological change. (For a further description see the OECD manual for patent studies OECD 1994b). Counting patents is directly related to the technological and industrial effects of R&D and may therefore also be employed as an indicator of relevance. Furthermore, special analysis may be developed on the relationship between science and technology. The extent to which scientific knowledge is applied in the innovation process may be indicated by the citations in the patent records. Thus such studies may be used for describing the science-technology relationship.

However, there are several limitations with using patents as R&D output indicators. It is important to realise that a wide range of innovations do not lead to patenting. Furthermore there are large differences between fields and nations when it comes to the tradition for patenting innovations. There will also be large variations in the technological or commercial importance between individual patents. Major methodological problems also arise because the sample of patents is quite often rather small. This is particularly poignant when analysing small countries or minor technological areas. In these cases the samples may be too small for meaningful statements (see Schmoch 1997). Because of this patents have obvious limitations as science and technology indicators and it may often be advisable to support such studies with additional information.

Innovation surveys and patent studies provide a wide range of information and indicators on innovation activity. Such indicators may be combined with other types of data. A combination of different indicators may for example show the effectiveness of the impact of firms investments in R&D, that is input-output measures like the amount of profit attributable to R&D divided by the investment in R&D (see part 1.4.1) Furthermore, several additional indicators in the technology area exist, for example on high technology products and the technological balance of payments (TBS). However, these kind of indicators are not described further here.

1.2.4 Other indicators

In addition to knowledge production, education is a main objective in most academic institutions. Thus, the number of graduates (including bachelor degree recipients and postgraduates) and Ph.D. degrees is often included as an output indicator at the institutional or national levels. These numbers may be related to the total number of students, or to the number of staff to indicate productivity. A particular manual is developed for indicators on human resources for science and technology (HRST). The concept of 'HRST' is much wider than R&D personnel, and indicators on HRST may yield additional information on results in terms of 'human competence' (compare the 'Canberra Manual' OECD 1994b).

1.3 Indicators of scientific impact

Output indicators such as publication counts do only give an indication of the volume of research, for example within a country. Quantity and quality are, however, not necessarily related. Additional information is needed to assess the impact, quality and the extent to which scientific activity may contribute to scientific progress. Obviously, this is something which is difficult to assess. Prominent among the indicators that may yield some information on these aspects are citation indicators.

1.3.1 Citation indicators

In research evaluation, citations have become a widely used measure of the impact of scientific publications. The basis for such studies is the number of citations to earlier scientific publications, as indicated in the reference lists of the research publications. The availability of computerised databases make it possible to conduct studies of citation patterns on aggregated levels.

The distribution of citations is extremely skewed. Most of the papers published get no or only one citation, whereas a relatively small number of documents receive a large part of the citations.[3] Generally, the number of citations to a particular article indicates how much the knowledge embodied in the paper is used by other scientists. Thus, the premise underlying citation indicators is that a research finding frequently cited by other researchers has had a greater impact on the research community than an infrequently cited paper. Often the number of citations is taken to be an indication of scientific quality. This assumption has, however, turned out to be rather problematic (see e.g. Seglen 1997). Authors cite one another for a variety of reasons. Quality is undoubtedly a highly complex matter and is not necessarily related to scientific impact.

Number of citations per paper

The most usual citation indicator at the macro level is probably the average number of citations per year within a country or scientific field. This is calculated by counting the number of citations to papers in a scientific field over a fixed time period (called "the citation window"). The number of citations divided by the number of papers receiving those citations yields the average number of citations per paper. If this number is above the world mean, it indicates that the research has a higher impact than the average.

There are great variations in the number of average citations per paper between different disciplines and in many cases between different subject areas within the same discipline. For example, within the same observation period, an average publication in the life sciences receives more citations than in the exact sciences or mathematics. Thus, a high citation rate in mathematics or theoretical physics may be considered moderate in the life sciences. Citation indicators must, therefore, be normalised if they are to compensate for these inequalities.

At a meso and macro level, appropriate citation-based indicators have proved to be valuable measures of research performance. Although this measure is somewhat

[3] Of the 65 mil. articles recorded in SCI from 1945 to 1988, the average number of citations per paper was 15. However, as many as 50 per cent did not get any citations at all and 28 per cent only got one citation (Garfield 1990). The number of highly cited papers with more than 50 citations constituted 0.8 per cent.

influenced by scientific profiles, as mentioned above, it affords an insight into the performance of national scientific research.

Generally, the methodological problems are more severe for smaller samples than for larger ones. For example, in micro-level studies, self-citations (citing own works leading to an increase in the number of citations with which a researcher will be credited) and citations received from within the research group, may influence the citation rates to a considerable extent and should therefore preferably be excluded. This problem is, nonetheless, rather minor if the volume of figures being analysed is high.

Attractivity Index
A particular indicator which is often used to measure national performance is the so-called attractivity index (see e.g. European Commission 1997). This relative indicator is based on comparing a publication's citation counts to the average number of citations accumulated by all publications in the ISI-covered journals in which it was published. These relative counts can be aggregated to the level of the scientific discipline to which the journal is assigned, which thus yields standardised measures removing journal- and/or discipline-dependent citation impact biases. The citation impact can be standardised according to two baselines: 1) on all international journals in the discipline, 2) to a narrowly defined baseline which is restricted to those journals in which one or more of a country's publications have appeared. More simply, such relative indicators can be calculated by dividing a country's share in citations given to its publications in the given field with the given field's share of citations received by all publications in the world (Schubert et al. 1988). Such indicators indicate whether the publications of a country attract more or fewer citations than anticipated on the basis of average citation rates. That is, if the country's citation impact is above or below the world average.

Several additional indicators involving citation counts exists, for example expected citation rates, number of high-impact papers, journal impact factors, etc. For a further description of such indicators see e.g. Okubu 1997.

Although citation analyses are gaining increasing popularity as a tool for the comparative assessment of researchers, research institutions, and countries, several problems are connected with such studies. As indicated above, there are serious methodological problems related to the construction of citation indicators. At a fundamental level the meaning of a citation count is not always clear. Therefore, such indicators should be used cautiously, particularly in evaluation exercises.

1.3.2 Prizes and academic distinctions

Prizes and other kinds of academic distinctions are awarded to researchers for eminent scientific work. The most prominent, is the Nobel Prize, although prizes exist on all levels of the research system and in a large area of scientific disciplines. Prestigious prizes, such as the Nobel Prize, are awarded to generally recognised research. New discoveries are rarely awarded.

Sometimes the number of prizes is used as an indicator. For example, the number of Nobel Prizes is included as an indicator in the EU report on science and technology indicators (see European Commission 1994; European Commission 1997). Typically, the number of Nobel prizes is employed as a cumulative measure of the national scientific level. Nevertheless, prizes are awarded on the basis of individual researchers' performance. A main question is, therefore, whether there is any significant relation between these relatively small number of researchers performances and the vitality and impact of research at a national level.

1.3.3 Indicators from peer evaluations

Peer evaluations are often considered to be the alternative to the system of output indicators. Such evaluations typically involve a qualitative assessment of the research performance by foreign researchers within the same field. In addition to the assessment of research performance, elements considered in peer reviews may include: editorships of outstanding journals, awards, membership of prominent scientific societies, honorary degrees, guest professorships at renowned universities, funding by research councils and by other external sources, advisory committee membership, etc. This method of assessment is normally the one favoured by scientists. It should, however, be recognised that quantitative bibliometric indicators of scientific production also involve a certain element of peer review, since this indicator is based on number of publication in refereed journals.

However, these kinds of evaluations are also recognised as having weaknesses. For example, there might be problems finding neutral peers, and no peer will have perfect information and will therefore base such evaluations on limited or imperfect information. To overcome some of these problems, quantitative output indicators are sometimes used as input in the peer process. Especially in circumstances where peer reviewes lose their power, bibliometric indicators can be used to support decisions. This may ensure that the judgements do not get too "soft" nor too "hard". New indicators may be created on the basis of such mixed qualitative-quantitative evaluations, for example by giving marks to factors like quality, productivity, relevance and viability (see for example evaluations in the Netherlands e.g. VSNU 1996).

1.4 Effects of R&D

When analysing the output of R&D it is usual to distinguish between the immediate results and products of research and the effects of this research. Science and technology have had a large influence on the development of modern society. The importance of science for social, cultural, technological and economic development means that the study of effects of R&D covers a wide range of areas. While much attention has been directed to developing indicators of the results of R&D during recent years, the study of the effects of R&D is far less developed. However, an increasing need to legitimise science as a social activity in the public sphere means that the interest in such studies is growing.

Generally speaking, broader and empirically well founded theories on how science influences society do not exist. The complexity of the interactions between science and society is of an order that makes such general theories more or less unattainable. As indicated above, factors from a wide range of areas have to be taken into consideration, for example the influence on economic developments, employment, political decisions, values and public opinion, health, welfare, the environment, etc. Even if one does not intend to explore the total effects on society, but restrict oneself to certain specific effects, the methodical problems are severe. In particular, such effects may be difficult or impossible to quantify and in turn be causally related to specific scientific activities (the problem of “attribution”).

However, even though no general framework exists to approach this vast and complex topic, researchers from various disciplines have done work on certain aspects of the effects of science. Some of these approaches are described below.

1.4.1 Economic effects of R&D

Because of the increasing focus on science as a driving force for technological and economic development, the study of the economic effects of R&D has been an area of major interest. Undoubtedly, this is also the area which has been most developed. Among other things, this is due to the fact that economic effects may be possible to quantify and in turn be used as basis for the development of different kinds of indicators.

The major approach in this area has been studies on the effects of R&D on productivity. Such studies range from examinations of specific innovations and their particular effects, to examination of aggregated estimates of productivity growth as a result of all R&D performed. Furthermore, such studies may focus on particular types of research, such as academic, government supported, and private, and on types of economic effects, such as cost reduction and quality improvement to the organisation performing the research, as well as spillover effects that benefit those who did not pay for the research.

The literature on this topic is vast. To mention a few examples, Terleckyj divided the effects of private R&D into two groups: 1) Direct increases in the productivity of industries conducting the privately financed R&D and 2) Indirect increase in the productivity of industries purchasing intermediate inputs from the industries conducting the privately financed R&D (Terleckyj 1980). In another study, Levy and Terleckyj examined government-financed R&D and observed that it had the effect of stimulating additional private R&D investments (Levy & Terleckyj 1982). Leyden and Link (1991) observed similar complementarity between government and private R&D.

Generally, economic analyses of R&D investments have concluded that R&D offers high returns in terms of higher productivity, although there are differences in the estimates of the exact level of returns (National Science Board 1996). This does not mean that every research project has a high, or even positive rate of return. Because results of research cannot always be predicted and often require a long time to develop, individual R&D investment carries an element of risk and, in many cases, R&D may not be the most profitable investment a firm could make. For example, economists have also observed equally high or sometimes greater rates of return for other types of investment by firms, e.g. the enhancement of productivity through the acquisition of new machinery, advertising, etc. (National Science Board 1996). The best type of investment would thus depend on the circumstances particular to a firm. At a national level, however, the average high rates of return to R&D observed by economists may serve as evidence of the importance of scientific research to the national economy.

1.4.2 Social effects of R&D

Despite the strong focus on the importance of science for technological and economic development, a large part of the research in science does not yield results that can be related to such factors. Nevertheless, the effects and utility of this research on society may be of tremendous importance. This will particularly be the case for academic research in medicine, the social sciences and the humanities, but also important in parts of the natural sciences. Despite this, analyses of the utility of research that cannot be measured in economic terms have been ignored to a considerable extent. The main reason for this is the problems mentioned above connected with identification of social effects and the difficulties with developing indicators of such relevance. The only viable method will often be a delimited and qualitative approach based on case studies. Below, some examples that have been used to study such effects in a quantitative manner are described. As will be seen, these studies have a rather tentative character and are not systematised to any further extent.

Generally, science will have both intended and unintended societal effects or relevance. Furthermore, these effects may be of a positive and negative nature. Some kinds of influences are revolutionary and lead to transformations in a relatively short period of time (e.g. new technology and important biomedical discoveries). Other kinds of influence may be related to gradual changes in thoughts, attitudes and cultural understanding (Dunn, Holzner, Hegedus, & Shahidullah 1988: 345). Generally, it may be useful to distinguish between “instrumental use” and effects in terms of “enlightenment”. Instrumental use may be a contribution to specific objectives or support for a particular decision, for example by decision-makers (Rich 1991). Enlightenment refers to the uptake of research results which have an impact on attitudes and behaviour.

Because applied science is often directed towards the solution of specific problems, such research may be of particular relevance to society. Applied research may also be easier to identify and connect to specific effects. However basic research may contribute to the solution of societal problems in the long run. Furthermore, the spillovers, or ripple effects, of basic research have often yielded high returns to subsequent applied research. Consequently, the overall net benefits of basic research to society may be quite high.

One new approach to the study of the societal effects of R&D was presented in the last indicator report by National Science Foundation (National Science Board 1998). Here, a new chapter was included on the economic and social significance of information technologies. The explosive growth in information technology has been one of the most dramatic effects of recent scientific and technological research. For example, this revolution has had major impacts on production and employment. Indicators of information technologies may thus illuminate one important aspect of the societal effects of R&D. Even though much research already has been done on this issue, it has not been approached in the context of R&D output indicators.

One attempt to study the transfer of scientific knowledge to a non-scientific public in a bibliometric way was done by Nederhof and Meijer (Nederhof & Meijer 1995). In some disciplines a large part of the research may be devoted to external knowledge transfer, that is, transfer of research results to primarily a non-scientific or non-scholarly public. For example, it is argued that particularly in the humanities a large part (sometimes more than 50 per cent) of the research output may be directed to such an audience. In more science-like fields such as experimental psychology, normally a more modest percentage of the output is directed to a non-scientific public (Nederhof & Meijer 1995). The study by Nederhof et al. focused on the importance of trade journals in the transfer of agricultural knowledge in the Netherlands. Because trade journals are rated by farmers as the most important source of knowledge in their occupation, such a study may indicate the possible utility of agricultural research. Thus, although limited in scope, this may be an example of a possible way to study the dissemination and uses of scientific knowledge in society.

Medicine is one area of research which most obviously has social effects and utilities. In a study by Grant et al. the impact of biomedical science in improvement in health care was demonstrated (Grant et al. 1998). Using clinical guidelines, that is a series of recommendations which, when followed, will lead to improvement in health, they proposed a new method to quantify the progress of knowledge from biomedical research to clinical practice. Bibliometric techniques were used to assess the link between scientific research and guidelines on acute low back pain, asthma and angina. By collecting the bibliographic details of the evidence cited in the guidelines, it was possible to identify what kind of research may be particularly important to such health care. In this way the study demonstrated that scientific research plays an important role in the creation of knowledge that underpins clinical guidelines.

One area were systematic surveys and indicators actually have been developed is the study of the "public understanding of science and technology" (PUST). Among others, such studies focus on the public interest in science and technology, the knowledge of science and technology issues, and the attitudes towards science and technology. During the last 20-30 years a relatively standardised set of indicators has been established on these issues. Pioneering work was implemented by the National Science Foundation, and indicators on PUST have been included in the NSF's "Science and Engineering Indicator Report" for several years . The possibilities for international comparisons were extended through PUST surveys that the EU implemented in 1989 and 1992 (EUROBAROMETER). Several S&T indicator reports now include such indicators (see e.g. European Commission 1994; Netherlands Observatory of Science and Technology 1996; Statistiska centralbyrån 1996). Although surveys on PUST have a somewhat different focus than the "effects of R&D" they may, nevertheless, yield important information on e.g. the influence of scientific discoveries on public attitudes. For example, science education at all levels is important for a better understanding of environmental issues, medical research issues, and other areas.

2 The uses of S&T output indicators in science policy

On the basis of the general overview presented in Chapter 1 we will now look more closely at the uses of S&T output indicators in science policy. Such indicators have to an increasing extent been employed in science policy processes and decision making. This has in particular been the case for bibliometric indicators related to academic research. In the area of technology and technology policy, other kinds of indicators have been required. However, our presentation below will mainly focus on the uses of bibliometric indicators in science policy processes. This does not imply that other kinds of output indicators have proved to be of little value in policy contexts. However, it reflects the fact that bibliometrics has been one of the main area of interest for science policy purposes. The presentation below focuses on the use of bibliometric indicators related to different levels of the research system. In addition some examples from practical applications in different countries are described.

2.1 Introduction

Generally, output indicators can have a number of objectives in science policy, monitoring and evaluations being the most important. In the context of monitoring, indicators are employed as tools for tracking developments and trends in the S&T area. For example, they may indicate characteristic features of a given S&T system and its infrastructure. A typical product that contains such information is S&T indicator reports.

The main area of uses of output indicators in general, and bibliometric indicators in particular, has been related to different kinds of research evaluations. Traditionally, peer review has been the standard method for evaluating research. Issues such as scientific impact, productivity, the level of research, strengths and weaknesses, potentials for future development, and the contributions of research to other academic fields and disciplines are typically addressed in such evaluations. The last 10 years bibliometric indicators have increasingly been used to support such peer-review processes. This is based on an awareness that decision-making might be informed by more systematic data on past performance within a discipline, institution or research group. Furthermore, with increasingly sophisticated ways of measuring the output of science, the relevance of these tools to policy makers has increased.

In this way qualitative peer review and quantiatative output indicators may be complementary. The strength of output indicators is due to their simplicity, the relatively easy access to relevant data and their (apparent) objectivity compared to classical peer review. There is a weakness due to the uncertainties of the relation between indicators and quality. That is, the issue of quality is too complex to be grasped adequately by quantitative methods.

In addition to input in peer processes, output indicators may be used in strategic analyses and in supporting different kind of decision-making processes. Typical questions are (Van Steen 1995):
– How does our country perform in S&T compared to other countries?
– Is the level of university research of a sufficiently high standard?
– To what extent do public research institutes interact with private enterprises?
– Is the innovative potential of those enterprises competitive enough?

Indicators can provide (partial) answers to these questions. In addition to giving insight into the performance of the S&T system, they may also have a function in respect to accountability and in justifying the way the R&D budget is spent.

With respect to policy-making, a distinction can be made between "direct" and "indirect" effects of output indicators. An example of a direct effect is when a policy maker refers, in his or her decisions or statements, explicitly to specific results or conclusions from the policy studies. Indirect effects occur when results from policy studies are used in the policy debate to raise relevant questions, clarify concepts, question assumptions or to substantiate impressions (OECD 1997b). Because indicators may help in clarifying policy debates and enhance the rational foundation of decision-making processes, they now have an increasingly powerful role in the context of science policy.

The Dutch have been particularly active in this area. At a conference on S&T indicators in 1988, it was claimed that the Netherlands and the United Kingdom were leading the development of S&T indicators (Nederhof & van Raan 1988: 197). This tendency has continued, as indicated by the examples below. At the same time there are considerable variations between different countries. Some countries have not been active in the area of R&D output indicators or tend to limit the use of such indicators. Still there is much work going on in this area. For example, the OECD organised a conference in Washington in 1997 called "Use of science and technology indicators for decisionmaking and priority setting".

There are, however, several obvious limitations connected with using output indicators in science policy. As seen in Chapter 1, a wide range of methodological problems are involved in bibliometric data processing. For example, in evaluations of social sciences and humanities, bibliometric indicators may be of limited value because of problems with database coverage (although bibliometric indicators have been used for research evaluation purposes even in these fields). Furthermore, if the importance of quantitative, bibliometric studies is overemphasised, this may possibly lead to a certain neglect of other aspects, e.g. education and research training. There are also several examples of policy makers misusing bibliometric data or using them in opportunistic ways. This is perhaps due to a lack of knowledge on how bibliometric indicators should be interpreted and on their limitations.

We will now look more closely at examples from uses of R&D output indicators in evaluations at different levels in the research system. At the first level these evaluations can focus on an entire scientific discipline. Secondly, they can concern larger research groups, laboratories, and institutions such as universities. At the third level, evaluations can focus on the work of individual researchers. Some examples from different countries are described below. This is not meant as a complete overview of uses of S&T indicators in science policy. Rather, the examples are illustrative and show different functions such indicators may have.

2.2 Evaluation of disciplines

At the highest level, R&D output indicators can involve aggregated data on a national level, for example on whole disciplines and internal sectors. This provides a basis for national benchmarkings and international comparisons.

In many countries such indicators are now used as input to national field evaluations. This may also involve more disagreggated indicators showing data for specific institutions and departments within the field evaluated. The United Kingdom and the Netherlands have in particular been active in applying such indicators in evaluation procedures.

2.2.1 The British evaluation system

Since 1986 a model based on R&D output indicators has been employed in evaluations of the British university sector (the UK Research Assessment Exercise). Every four year all subjects are evaluated (69 categories). The evaluation is organised by the Higher Education Funding Council and is implemented by national panels for each subject (Edwards 1998). Input to the peer process is data on publications, external support (e.g. grants) and statements of policy. Bibliometric data such as numbers of publications and citations are in this way important indicators. The outcome is a quality assessment and evaluation marks on a 7-point scale.

The results of the evaluations are directly employed in external policy decisions on the allocation of research resources. In Britain there are two main sources of government funding in the research system: block grants which are provided to support infrastructure and the basic capabilities in universities to carry out research and teaching, and the "science budget" which mainly pays for individual research projects and postgraduate training (OECD 1997b). In both cases funding are based on evaluations: the block grants on the results of the assessments of individual departments in the universities, and the science budget on the basis of reviews of research applications. The base funding for the individual departments is determined on the basis of volume (number of researchers) and the evaluation score. In this way the results of the assessment have direct consequences for the resource situation within a department.

The Research Assessment Exercise was introduced to consolidate the existing division between institutions with much research and institutions with little or no research. The allocation method has also resulted in such a division: a small group of universities is highly concentrated on research, a middle group with both teaching and research and a large part almost completely concentrating on education (for example, 21 of a total of 170 universities have only 22% of the total student population but receive 64% of the total research resources (Westergaard 1995)).

The allocation method has, however, turned out to have several negative consequences. For example, historically strong research institutions are favoured at the expense of promising research groups, employment of researchers is based on short-sighted gains, "headhunting" and personnel dispositions, it undervalues multidisciplinary research and stimulates publication patterns on quantitative more than qualitative criteria (Hansen & Jørgensen 1995). In addition, the evaluation is costly and time consuming. Nevertheless, the rating system is expected to be maintained in future allocation processes (Edwards 1998).

2.2.2 The Dutch evaluation system

In the middle of the 80s, bibliometric indicators were used for the first time as input to policy decisions in merging and closing university faculties in the Netherlands. (Van Steen 1995; Van Steen & Eijffinger 1998). This caused a lot of discussion about the value of such indicators. The Ministry of Education and Science, however, became interested in bibliometrics as an instrument for measuring aspects of R&D. Between 1987 and 1992 a five-year research programme was financed at the Centre for Science and Technology Studies (CWTS) at the University of Leiden. The aim was to do basic research on bibliometric indicators. A new five-year programme (1992-1997) continued this endeavour. This time the aim was to do more policy-oriented research on S&T indicators.

On the basis of the research within CWTS, the Netherlands now has a leading role in the field of bibliometrics. R&D output indicators have also been used widely in respect to evaluations and decisions-making in Dutch science policy. In contrast to the UK, the evaluations have had a diagnostic function and have not been coupled to resource allocations.

At the level of disciplines, bibliometric indicators are used as input to regular field evaluations (see e.g. VSNU 1996). The Association of Universities in the Netherlands (VSNU) has set up a system for such research evaluations. The system covers all academic research, is disciplinary oriented, and addresses research at the programme level. Every five years international peers assess the quality of specific disciplines. In this process, bibliometric indicators are used variably as input (OECD 1997b). The resulting system is intended to be used as a point of departure in a dialogue between the evaluated persons and units and the evaluators about strengths and weaknesses as well as potentialities. Apart from being a direct feedback to research managers and researchers, the system is also intended to strengthen a quality oriented development of institutional management. The evaluations involve four aspects:
– *Scientific quality*: Involves assessment of the quality and international prominence of the research group. For example, the quality of dissertations and publications, the originality and coherence of research and the contribution to the development of the discipline are assessed.
– *Productivity*: Involves indicators such as the number of dissertations, the number of publications in scientific journals and books, the number of patents, and the number of invited lectures. The research output is related to human and material input resources.
– *Scientific relevance:* Here the research issues and approaches are weighted. In particular the importance of the research for the advancement of knowledge within the discipline (scientific relevance), and significance with regard to societal/technological impact (e.g. possible impact and application in future technologies) are addressed.
– *Long-term viability*: This aspect is assessed on the basis of submitted plans and ideas for future research, and the availability of personnel and facilities. Furthermore, the direction the research is taking and competitive strengths that may depend on factors of scale and the scientific infrastructure available are evaluated.

On every aspect a mark from 1-5 is given, where 5 is the best (that is: poor/unsatisfactory/satisfactory/good/excellent). As an integrated part of the evaluation, written comments on all four aspects are added. The results of the evaluation are presented in a report (e.g. VSNU 1996).

The bibliometric analyses and indicators used as elements on which decisions are based include for example (Van Raan 1993):

– The number of papers in international journals (covered by ISI) and trend in scientific productivity.
– The number of external citations (self-citations and citations by the group itself excluded) received in the first three years after publication.
– The number of citations per paper, compared with the same ratio for an average paper (world-wide) in the set of journals used by the group. This is regarded as a reasonable indicator of the international esteem of the research group.

VSNU has concluded that the bibliometric indicators have been a valuable tool in these research evaluations. A combination of such indicators with qualitative approaches based on written and oral information has increased the reliability of the assessments. However, bibliometric analysis cannot replace judgements by peers. On the other hand, peer judgement alone will not give sufficient information on important aspects of research productivity and on the impact of research activities (Van Raan 1993).

2. 3 Evaluations of institutions and departments

As described above, discipline evaluations may include evaluations on the level of individual institutions and departments. There are also several examples on the use of S&T output indicators in individual evaluations of particular institutions and departments.

In the early 1980s Martin and Irvine performed a particularly influential study (Martin & Irvine 1983). In their study of radio astronomy a variety of data was applied, ranging from bibliometric indicators (number of scientific papers, total citations to recent papers, citations per paper, and number of highly cited papers) to structured tapping of peers' judgements in interviews. The information was used to support their assessments of big telescope institutes. Through this study they demonstrated that research output could be measured in a manner acceptable to the scientific community itself.

The last fifteen years there have been several studies involving S&T output indicators in the evaluation of universities. As one example, in 1990 the University of Ghent was the first Belgian university that decided to conduct systematic evaluation of its research performance on the basis of bibliometric indicators. The study involved an assessment of the research activities at the faculties of medicine and science during the 80s. The methodology applied involved the combination of bibliometric analyses and a validation by the scientists involved.

Similar studies have been performed for the assessments of research performance in for example universities in the Flanders and the Netherlands (OECD 1997b).

Typically, these studies have involved the use of bibliometric indicators as an input source to peer-based evaluations. In respect to a distinction between the direct and indirect effects of such policy-relevant studies, the effects here were mostly indirect. That is, no explicit use has been made of the results obtained to for example allocate funds. On the other hand, these studies have provided useful information to evaluators and policy makers in Flanders and the Netherlands (OECD 1997b). For example, in respect to a following debate about the creation of "centres of excellence" and a stronger concentration of research capacity on a limited number of topics, the studies formed valuable background material.

2.4 Evaluation of individual researchers/ research proposals

Funding committiees have traditionally relied upon evaluations by expert referees in making decisions on grant applications. In an explorative study by the Wellcome Trust in England, the benefit of bibliometric indicators as an aid to such decision-making processes was investigated (Lewison, Cottrell, & Dixon 1998). The study involved different indicators like the volume of production (number of pages as well as number of papers) and its influence as measured by the impact factor of the journals and the number of citations received by individual articles. The bibliometric indicators based on the publication track records of the applicants were then compared with the panel's funding decision in order to see how well the two correlated.

The study showed that the funding decisions usually were consistent with the bibliometric analysis of an applicant's work. However, some applicants with strong track records were refused because of the design of the proposed research programme. Generally, the panel members found the bibliometric indicators helpful in informing their discussions. The indicator of greatest utility was the number of articles among the most cited 25% (top quartile) of the comparison group. The method showed to be the most useful for the appraisal of senior scientists with a large number of papers, but less so for scientists who are still early in their careers.

Thus, also at the level of individual research proposals, S&T output indicators have been showed to be useful. However, the study recognised that because the interpretation of bibliometric indicators requires a great deal of specialised knowledge, they should be used with great caution.

3 S&T output indicators in Norway: data sources, research and applications

In Norway, as in most other countries, activity in the field of S&T indicators has traditionally been concentrated on input resources. The work on S&T output indicators has been relatively limited. However, the last decade the focus has increasingly been shifting towards output indicators. This research has covered both bibliometric and technological output indicators. In this chapter we briefly describe some of the work within this area and also how S&T output indicators have been used in science policy. The chapter is not meant as a complete overview. For further and more detailed descriptions we refer to works produced by the involved institutions.

The strategy department at the Research Council of Norway has a particular responsibility for the knowledge base for research policy in Norway, and it is also the main source in financing work on science and technology indicators. The institutions particularly involved in these activities are Statistics Norway (SSB), the STEP group, and the Norwegian Institute for Studies in Research and Higher Education (NIFU). As one result of this work, a new report on science and technology indicators covering a wide range of indicators has been produced (Norges forskningsråd 1997). The report represents a collaboration between the Research Council of Norway, SSB, the STEP group, and NIFU. Similar reports will be published every second year. Although the report also included indicators on the output side, this area was not particularly well developed, due to a deficiency of relevant data and analyses.

3.1 Bibliometric indicators

In the field of bibliometrics, there have been some activities the last ten years at the Norwegian Institute for Studies in Research and Higher Education. Bibliometrics has not had permanent funding at NIFU and has been performed mostly on an *ad hoc* basis on the commission from the Research Council of Norway and other national and Nordic agencies. In particular this work has included the production of bibliometric indicators for Norway and the Nordic countries, see e.g. (Sivertsen 1991; Sivertsen 1992; Olsen et al. 1994; Sivertsen 1993; Olsen & Kaloudis 1997b; Olsen 1998). Many of these analyses have been based on bibliometric products and databases from ISI, such as the National Science Indicators on Diskette (NSIOD) containing aggregated publication and citation counts for different fields of science, and the National Citation Reports, containing bibliometric data for all Norwegian publications in ISI-indexed journals.

More critical reviews on the use and interpretation of bibliometric data include works by P. O. Seglen (e.g. Seglen 1989; Seglen 1994; Seglen 1996; Seglen 1997a; Seglen 1997b). The research in the field of bibliometric and related areas has been relatively limited. Worth mentioning are the studies by S. Kyvik on productivity in academia (see e.g. Kyvik 1990; Kyvik 1991; Kyvik 1993, Olsen & Kyvik 1993).

Compared to countries like the Netherlands and Great Britain, the use of bibliometric indicators in research policy has been relatively sparse in Norway. However, the last years such data has sometimes been used in respect to research evaluations. For example, in the field evaluations of Norwegian geosciences and chemistry, bibliometric indicators were produced to assist the peer evaluation processes (see Olsen & Kaloudis 1997a, Kaloudis & Olsen 1998).

In the future, the activity in the field related to output indicators, and bibliometric indicators in particular, will be strengthened through a new strategic research programme at NIFU. This programme running in the period 1999-2001, will focus on profiling the output of Norwegian research.

In 1993 a working group on statistics, information, and documentation on R&D and R&D-related activities was established under the direction of the Research Council of Norway. Among other things, this group addressed how the knowledge base for research policy could be further developed. The group consisted of members from different institutions (the Research Council of Norway, the Norwegian Institute for Studies in Research and Higher Education, Statistics Norway, the Norwegian Social Science Data Services, and the STEP group). The work of the group resulted in a report in which available sources were reviewed and recommendations for strengthening work in the field were proposed (Norges forskningsråd 1994).

Among its recommendations was to further develop the systems for project and output documentation established by the earlier Research Councils and several institutions into a standardised and national system for the higher education sector. How such a system could be established has now been on the agenda for several years, and is in the process of being planned through cooperation between the Research Council and the higher education sector. However, it will probably take many years before it is functioning and even longer before it will be suitable for output analyses.

Today the system for project and output documentation that has been most extensively developed is the FORSKDOK system developed by BIBSYS. The institutions in the higher education sector have in varying degrees applied this system for documentation purposes. In particular the University of Bergen has used the system extensively, and it is said to have almost 100% coverage of the publication output of this university. The Norwegian University of Science and Technology (NTNU) is also reported to have fairly good coverage. Furthermore, the Norwegian Social Science Data Services (NSD) is responsible for developing the infrastructure for integrating the institutional databases (or institutional uses of FORSKDOK) into a national database for research documentation. This system will include a broad range of outputs of research activities.

3.2 Technology and innovation indicators

During the 80s increasing interest was directed towards the study of technology and innovation. This particularly took place within a group at the Norwegian Computing Centre (now STEP group), but also at Statistics Norway (SSB).
The STEP group (Studies in technology, innovation, and economic policy) and SSB have been responsible for most of the work within the area of innovation and technology indicators in Norway. Several databases relevant for analyses of technological and economic output of R&D and innovations are located at SSB, among these are: industry statistics, R&D statistics (business enterprise sector), accounts statistics and statistics on technological balance of payments. In addition SSB is responsible for implementing the innovation surveys. This means that Norway, in respect to data, is in a relatively good position for doing analyses in this area. For example, there have been several surveys involving linking different data sources. While SSB is the principal institution responsible for data collection, the STEP group has been in charge of most of the research in the area of technology and innovation indicators.

One example is the coupling of R&D statistics with other economic statistics at the enterprise level. In this way the development of profit, productivity, size, etc., in relation to R&D investments has been studied. Such analyses have furthermore showed differences between enterprises in different trades and between R&D enterprises within the same trade. In this way such studies have illuminated important aspects related to the effects of R&D.

In respect to S&T output indicators, the innovation surveys are particularly important. The indicators are based on the *Oslo Manual* and were first used in the innovation survey of 1992. A sample of the indicators was included in the R&D survey for 1995. The second innovation survey for 1997 has now been completed. Through these surveys it has been possible to look at the distribution of innovation activites among the enterprises, the distribution of innovation expenses, the results of innovations as measured by share of new or changed products in trade, in addition to the relation between innovations and economic profits. The results of last innovation survey (CIS 2) have not yet been published. The results of the first two surveys have been published in reports such as the Norwegian indicator report (Norges forskningsråd 1997).

The STEP group has continued the work on S&T output indicators in this area, especially around the so-called IDEA project ("Indicators and Data for European Analysis") for the European Commission.

Research on patenting activities is increasing internationally. In Norway the activity is rather sparse. Data on the patenting activities have been included in the Norwegian Indicator Report and there are some additional studies (e.g. Iversen & Kaloudis 1998).

3.3 Other indicators and further developments

In addition to the data sources mentioned, there are some other sources that are suitable for output analyses. For example, the key indicator survey for the institute sector also includes information on published output, not only international journal papers, but also all kinds of scientific literature.

Concerning academic science, data on education and the production of researchers may be used as indicators. For example, there are databases covering the number of candidates and Ph.D. degrees. Such data may be included as output indicators on an institutional or national level showing the production of new researchers and, indirectly of new competence.

The area of social effects (non-economic) of R&D has not received much attention in Norway. Although policy makers demand such knowledge, few studies have adressed the issue. One example involving a qualitative case study is the investigation of the policy impact of the acid rain research programme in Norway, carried out in the period 1972-1980. A main objective behind the acid rain programme was to create a knowledge base for political actions (Roll-Hansen & Hestmark 1990). The programme represented an example of how science can function in advising political decision processes, although the study of this programme did not primarily concern the level of output or effect.

As we have seen there are several activities in the area of R&D output indicators in Norway. It is expected that this area will become an increasingly strategically importance in the future. For example, the R&D system in Norway is continously being meet by demands for evaluation and result documentation. The need for documentation is related to all aspects of the R&D system, for example in planning, shaping of means, and in relation to difficult policy choices and priorities. Indicators on the outputs and effects of R&D may provide part of the knowledge required.

Increased efforts are required for further advances in the field of R&D output indicators in Norway. This may involve more research, better use and combination of existing data and databases, and the application of new data sources. Concerning the latter, the new system for research documentation may be of particular importance. When functioning, this system will provide systematic information on other types of results than articles in scientific journals (reports, books, popular science, etc.) and in this way supply standard bibliometric indicators based on ISI products.

One particular challenge will be to develop a system of categories (result profiles) that manage to intercept a wider range of knowledge products, used for example for the purpose of evaluation of the productivity and effects of different kinds of research activities. Increased knowledge in this area will be important for example for identifying factors that influence the effectiveness of various kinds of policy measures, organisational frameworks, and mechanisms for allocating resources.

Concluding remarks

This report has attempted to give a description of different kinds of S&T output indicators and related studies in an international and national context. On the basis of the underlying review, it is tempting to say that no substantially new kinds of output indicators have been developed in the last years. The indicators available are to a large extent based on products of R&D that traditionally have been applied as output indicators, like publications and patents. This may indicate that one has reached a limit when it comes to finding entities that are suitable for quantitative output analyses. At the same time, there has been important progress in new ways of applying the data available for indicator purposes, for example by co-word and co-citation analyses.

Despite the advances in the development of S&T indicators the last few years, many shortcomings still exist. For example, the traditional indicators have mostly focused on the natural sciences, medicine, and technology. Indicators covering the area of social science and humanities and societal effects of R&D are far less developed. Furthermore, such indicators tend to neglect "grey" literature like reports, which may have an important function particularly in applied research as well as in innovative development.

The science system has undergone many changes the last years. Some of these changes have created particular new needs in respect to indicators. For example, these changes have been described as an evolution from a "linear" to an "integrated" science and innovation system (Gabolde 1998). According to Gabolde, factors characterising these changes include: increasing globalisation, a society increasingly based on knowledge, increasing interaction between S&T development and socio-economic goals and challenges, the increasing cost of research itself and of its infrastructure in a context of increasingly tighter budgets. Such changes mean that new "systemic indicators" describing connections and interactivity, but also a wider range of output indicators are required.

In addition there is often a gap between the information policy makers want and the available indicators. That is, the needs of policy makers are not satisfied. Thus one challenge is to design and produce indicators that mirror the diversity of decision-making needs, e.g. on factors such as the scope of research, related to problems of specific national interest, criteria of societal relevance, etc. Another challenge is to develop a better way to introduce new knowledge and indicators into decision-making processes (compare Barré 1997). There are also more specific challenges related to, for example, achieving a greater visibility of scientific research within society. Here, traditional indicators are not particularly suitable.

Some of these shortcomings may be reduced by a more extensive use of the existing databases. Several institutions and organisations are involved in the work on further development of S&T indicators. For example, working groups of the OECD countries meet regularly to develop methods for collecting and applying internationally comparable statistics and indicators. The OECD has now selected six priority areas for the development of a new generation of indicators (new S&T indicators for a knowledge based economy): the mobility of human resources, patents, innovative and absorptive capabilities of firms, internationalisation of R&D, government support to industrial R&D, and information and communication technologies (ICT).

Despite the possibilities for further developments within this area, the limitations related to output indicators should be carefully recognised. The explanatory power of such indicators has obvious restrictions. One fundamental limitation is related to the fact that indicators as such only focus on measurable aspects of the science system, that is they are limited to aspects of results and effects that can be quantified. However, many aspects of the results of R&D cannot be measured in a meaningful way or captured by quantitative indicators. For example, knowledge production is one principal outcome of research. When using publications as an output indicator, this dimension is neglected. In particular one does not address the core point: the content of the research. Therefore, one should not overestimate the information R&D output indicators actually can provide and the value they may have in science policy and as tools for analysing research. Because indicators can only give a partial picture of the actual outputs and effects of R&D, more qualitative approaches based on case studies may sometimes yield important additional information.

A research policy system increasingly based on output indicators may have several undesired consequences. As we have seen for Great Britain, if applied to research allocations, the publication and citation behaviour of researchers may change in very unfortunate ways. Furthermore, an increased use of output indicators may create undesired differences when it comes to attention from policy makers. In particular the focus on non-measurable aspects can be reduced. Because many aspects of the results of R&D are difficult to quantify, there is a tendency to use "easy data" (readily available, but not necessarily characteristic for the aspects one would like to monitor). This is an obvious problem with many indicator-based approaches to R&D today. Furthermore, there are several examples of misuse or abuse of indicators, especially among policy makers. In particular, the restrictions and limitations of the indicators are not always taken into consideration when applied in research policy.

Among researchers, there has often been a scepticism towards the use of S&T output indicators in science policy. This may particularly be related to a suspicion that policy makers will take control over decisions away from the research community itself. At the same time, the more sophisticated examples of use of such indicators generally seem to have received an acceptance within the scientific community. This emphasises the necessity of applying indicators with great care and that the limitations in the quality of the output indicators should be clearly addressed when used in science policy. It is therefore important that indicator research remains part of the quantitative study of science and technology, developing R&D indicators should not be seen as pure consulting service work.

In any case, output indicators can be expected to become increasingly important in future science policy. Despite the problems and weaknesses related to R&D output indicators, they nevertheless provide interesting and valuable information on the structure and function of the R&D system. Thus careful use of S&T indicators, with full knowledge of their limitations, is generally preferred rather than none at all. Since individual indicators can only give a partial picture, and because they have different strengths, weaknesses and problems of validity, a combination of different indicators is often required, for example by producing output profiles. Indicators may be especially valuable when they are fed into the mechanism traditionally employed in science policy, namely peer review (e.g. Irvine et al. 1987).

References

Aksnes, D. W., Olsen, T. B., & Seglen, P. O. (1998): Inadequacy of a journal-based research field delineation. Incomplete recovery of Norwegian microbiology articles in ISI's Microbiology field. Paper submitted for the 7th ISSI Conference, Mexico July 1999.

Barré, R. (1997): The European Perspective on S&T Indicators, *Scientometrics,* 38(1), 57-70.

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.

Bruin, R. E. D., & Moed, H. F. (1993): Delimitation of scientific subfields using cognitive words from corporate addresses in scientific publications, *Scientometrics,* 26(1), 65-80.

Callon, M., Law, J., & Rip, A. (1986): *Mapping the Dynamics of Science and Technology,* London: MacMillan.

Carpenter, M. P., Gibb, F., Harris, M., Irvine, J., Martin, B. R., & Narin, F. (1988): Bibliometric Profiles for British Academic Institutions: An Experiment to Develop Research Output Indicators, *Scientometrics,* 14(3-4), 213-233.

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*, Amsterdam: Harwood Academic Publishers.

Dunn, W., Holzner, B., Hegedus, A. M., & Shahidullah, M. (1988): *Toward a Metric of Science Impact: Some Conceptual Explorations.* Paper presented at the conference Science and Technology Indicators. Their Use in Science Policy and Their Role in Science Studies, Leiden, 14-16 November 1988.

Edge, D. (1995): Reinventing the Wheel, in S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (eds.): *Handbook of Science and Technology Studies*, London: SAGE Publications.

Edwards, K. (1998): *Research assessment in Europe, with focus on the English system*, Lecture given at the University of Oslo, 20.03.98.

European Commision (1994): *The European Report on Science and Technology Indicators 1994,* Luxembourg.

European Commision (1997): *Second European Report on S&T Indicators 1997,* Luxembourg.

Gabolde, J. (1998): New challanges for indicators in science and technology policy-making: a European view, *Research Evaluation,* 7(2), 99-104.

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. (1997): The ISI Database: the journal selection process, *The ISI Essays, Internet.*

Grant, J., Cluzeau, F., Littlejohns, P., & Anderson, J. (1998): *Developing Indicators to Measure the Impact of Biomedical Research in Health Care.* Paper presented at the conference 5th International Conference on Science and Technology Indicators, Cambridge, 4-6 June 1998.

Hansen, H. F., & Jørgensen, B. H. (1995): *Styring af forskning: Kan forskningsindikatorer anvendes?,* Fredriksberg: Samfundslitteratur.

Irvine, J., & Martin, B. (1984): CERN: Past Performance and Future Prospects: II. The Scientific Performance of the CERN Accelerators, *Research Policy,* 13, 247-284.

Irvine, J., Martin, B. R., Abraham, J., & Peacock, T. (1987): Assessing basic research: Reappraisal and update of an evaluation of four radio astronomy observatories, *Research Policy,* 16, 213-227.

Iversen, E., & Kaloudis, A. (1998): Knowledge-bases and intercations in the Norwegian system: A patent-share and citation analysis, in K. Smith (ed.): *Mapping the Norwegian knowledge system*, Oslo: STEP-group.

Kaloudis, A., & Olsen, T. B. (1998): *Publication and Citation within Norwegian Geosciences. A bibliometric analysis,* Oslo: Norwegian Institute for Studies in Research and Higher Education.

Katz, S. J., & Hicks, D. (1998): *Indicators for Systems of Innovation,* Oslo: STEP group. (IDEA 12-98.)

Kyvik, S. (1990): Age and scientific productivity. Differences between fields of learning, *Higher Education,* 19, 37-55.

Kyvik, S. (1991): *Productivity in Academia. Scientific Publishing at Norwegian Universities,* Oslo: Universitetsforlaget.

Kyvik, S. (1993): Academic staff and scientific production, *Higher Education,* 5, 191-202.

Laredo, P., Mustar, P., Callon, M., Birac, A. M., & Fourest, B (1992): Defining the strategic profile of research labs: The "research compass card method" in Van Raan, A. F. J., Bruin, R. E., Moed, H. F., Nederhof, A. J., & Tijssen, R. W. J (eds.) *Science and technology in a policy context,* Leiden DSWO Press.

Leeuwen, T. N. v., 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 (CWTS, 96-09.)

Levy, D. M., & Terleckyj, N. E. (1982): *Effects of Government R&D on Private R&D Investment and Productivity.* Paper presented at the conference Annual Meetings of the Southern Economic Association, Atlanta, November 11 1982.

Lewison, G., Cottrell, R., & Dixon, D. (1998): *Bibliometric Indicators to Assist the Peer Review Process.* Paper presented at the conference Fifth International Conference on Science and Technology Indicators, Hinxton, Cambridge, June 4-6 1998.

Leyden, D. P., & Siegel, D. (1991): Why are Governmental R&D and Private R&D Complements?, *Economic Inquiry,* 29(2), 203-229.

Leydersdorff, L. (1995): *The Challange of Scientometrics: The Development, Measurment, and Self-orgainization of Scientific Communications,* Leiden: DSWO Press.

Martin, B. R. (1996): The use of multiple indicators in the assessment of basic research, *Scientometrics,* 36(3), 343-362.

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

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

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., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1987): *On the Measurement of Research Performance: The Use of Bibliometric Indicators,* Leiden: University of Leiden.

Moed, H. F., & Velde, J. G. M. v. d. (1993): *Bibliometric profiles of academic chemistry research in the Netherlands,* Leiden: Center for Science and Technology Studies (CWTS, 93-08.)

National Science Board (1996): *Science & Engineering Indicators - 1996,* Washington DC: U.S. Government Printing Office.

National Science Board (1998): *Science & Engineering Indicators - 1998,* Arlington: National Science Foundation.

Nederhof, A. J., Leeuwen, T. N. v., & Visser, M. S. (1997): *A Bibliometric Study of the Faculty of Mathematics and Natural Sciences, University of Groningen (1980-1995),* Leiden: Center for Science and Technology Studies (CWTS, 97-03).

Nederhof, A. J., & Meijer, R. F. (1995): Development of bibliometric indicators for utility of research to users in society: measurment of external knowledge transfer via publications in trade journals, *Scientometrics,* 32(1), 37-48.

Nederhof, A. J., & van Raan, A. F. J. (1988):*The Use and Development of Science and Technology Indicators for Policy Making in Five Countries: A Comparison.* Paper presented at the conference Science and Technology Indicators. Their Use in Science Policy and Their Role in Science Studies, Leiden, November 14-16 1988.

Netherlands Observatory of Science and Technology (1996): *1996 Science and Technology Indicators. Summary,* Leiden: HAVEKA bv and Karstens Drukkers bv.

Netherlands Observatory of Science and Technology (1998): *1998 Science and Technology Indicators. Summary,* Leiden: HAVEKA bv and Karstens Drukkers bv.

Nordisk Industrifond (1992): *Vitenskaps- og teknologiindikatorer for Norden – en artikkelsamling,* Informasjon 3/1992.

Nordisk Ministerråd (1996): *Vitenskaps- og teknologiindikatorer for Norden – en artikkelsamling,* København, Nord 1996:14.

Norges forskningsråd (1994): *Statistikk, informasjon og dokumentasjon om FoU og FoU-relatert virksomhet,* Oslo.

Norges forskningsråd (1997): *Det norske forskningssystemet - statistikk og indikatorer 1997,* Oslo.

OECD (1994a): *Frascati Manual 1993. The Measurment of Scientific and Technological Activities. Proposed Standard Practice for Surveys of Research and Experimental Development,* Paris: OECD.

OECD (1994b): *The Measurment of Scientific and Technological Activities: Using Patent Data as Science and Technology Indicators - Patent Manual*: OECD/GD 94/114.

OECD (1997a): *Oslo Manual. Proposed Guidelines for Collecting and Interpreting Technological Innovation Data,* Paris: OECD/Eurostat.

OECD (1997b): *The evaluation of scientific research: Selected experiences,* Paris: OECD.

Okubu, Y. (1997): *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples,* Paris: OECD. (STI Working Papers 1997/1.)

Olsen, T. B. & Kyvik, S. (1993): Publisering i den samfunnsvitenskapelige instituttsektor i Olsen, T. B. (red.): *Institutter for anvendt samfunnsforskning,* Oslo: NORAS.

Olsen, T. B., Hansen, H. F., Luukkonen, T., Persson, O. & Sivertsen G. (1994): *Nordisk forskning i internasjonal sammenheng,* København: Nordisk Ministerråd (TemaNord 1994:618).

Olsen, T. B. (1998): *Norsk forskning i internasjonale tidsskrifter. Sammenligning med andre land belyst ved bibliometriske makrodata.,* Oslo: Norsk institutt for studier av forskning og utdanning (NIFU). Rapport 1/98.

Olsen, T. B., & Kaloudis, A. (1997a): *Publication and Citation within Norwegian Chemical Research,* Oslo: NIFU. Skifterserie 12/97.

Olsen, T. B., & Kaloudis, A. (1997b): *Publisering og sitering innen medisinsk forskning. En bibliometrisk analyse,* Oslo: NIFU. Skiftserie 9/97.

Rich, R. F. (1991): Knowledge Creation, Diffusion and Utilization: Perspectives of the Founding Editor of Knowledge, *Knowledge: Creation, Diffusion, Utilization,* 12(3), 319-337.

Roll-Hansen, N., & Hestmark, G. (1990): *Miljøforskning mellom vitenskap og politikk,* Oslo: NAVFs utredningsinstitutt. (Melding 1990:2.)

Schmoch, U. (1997): Indicators and the relation between science and technology, *Scientometrics,* 38(1), 103-116.

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. (1989): From Bad to Worse: Evaluation by Journal Impact, *Trends Biochem Sci,* 14, 326-327.

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. (1996): Quantification of scientific article contents, *Scientometrics,* 35(3), 355-366.

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

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

Sirilli, G. (1998): *Innovation Indicators in Science and Technology Evaluation.* Paper presented at the conference Science and the Academic System in Transition. An International Expert Meeting on Evaluation, Vienna, July 3-5 1998.

Sivertsen, G. (1991): *Norsk forskning på den internasjonale arena. En artikkelsamling av 18 OECD-lands artikler og siteringer i Science Citation Index 1973-86,* Oslo: NAVFs utredningsinstituttt. Rapport 1/91.

Sivertsen, G. (1992): Bibliometriske vitenskapskapsindikatorer: *Vitenskaps- og teknologiindikatorer for Norden. En artikkelsamling,* Oslo: Nordisk Industrifond.

Sivertsen, G. (1993): *Nordisk samfunnsforskning i internasjonale tidsskrifter,* København, Nordisk Ministerråd (Nord 1993:5).

Smith, K., Ekeland, A., Iversen E., Kaloudis A., Patel, P., & Narula, R. (1998): *Understanding Science, Technology and Innovation Indicators: A Guide for Policymakers,* Oslo: STEP-group, Idea report 5/1998.

Statistiska centralbyrån (1996): *Vetenskaps- och teknologiindikatorer för Sverige 1996,* Halmstad: Bulls Tryckeriaktiebolag.

Terleckyj, N. (1980): Direct and Indirect Effects of Industrial Research and Development on the Productivity Growth of Industries, in J. W. Kindrick & B. N. Vaccara (eds.): *New Developments in Productivity Measurement and Analysis*, Chicago: University of Chicago Press.

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 Steen, J. (1995): Dutch indicators. S&T indicators in science policy: how can they matter?, *Research Evaluation,* 5(2), 161-166.

Van Steen, J., & Eijffinger, M. (1998): Evaluation practices of scientific research in the Netherlands, *Research Evaluation,* 7(2), 113-122.

VSNU (1996): *Quality Assessment of Research - Chemistry, past performance and future perspectives,* Utrecht: VSNU.

Westergaard, J. (1995): *Højreradikale spændinger i britisk universitetspolitikk,* København: Sociologisk instituts Sociologiske Rapportserie (1-95).