BIBLIOMETRICS AND INSTITUTIONAL EVALUATION

Jane M. Russell, *Senior Researcher, Centro Universitario de Investigaciones Bibliotecológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico*

Ronald Rousseau, *Professor, Department of Industrial Sciences and Technology, KHBO, Oostende, Belgium*

Keywords: Bibliometrics, Scientometrics, Research groups, Research evaluation, Research institutions, Research performance, Scientists, Peer review, Science and technology indicators, Scholarly journals, Journal Impact factors, Science policy, Citation analysis, Bibliographic databases, Science Citation Index.

Contents

Glossary

Bibliometrics: the application of quantitative methods to books and other communication media.
IF: **Journal Impact Factor**: basically a ratio between citations and citable items published in journals.
ISI: **Institute for Scientific Information**: Philadelphia-based company providing scholarly information services including the citation indexes and the Journal Citation Reports.
JCR: **Journal Citation Reports** : a publication of ISI providing bibliometric data for journals covered by their citation indexes including the journal impact factors.
Mainstream journals: generally used to refer to a group of highly cited and highly visible journals covered by the ISI citation indexes.
Peer review: evaluation of a scientist's work by his/her colleagues with similar scientific interests and rank.
SCI: **Science Citation Index**: a publication of ISI indicating who cites whom, used as a source of data for citation analysis and journal impact factors in science, technology and medicine.
Scientometrics: the study of the measurement of scientific and technological progress.

Summary

All evaluations are dependent on the availability of adequate and reliable data relating to the outcome of the activities under scrutiny. Literature-based or bibliometric indicators which quantify the production and use of bibliographic material, have been used extensively in the assessment of research performance. Their use is based on the assumption that the immediate purpose of research is to produce new knowledge and that publication is the primary form of output. Publication counts serve as an indicator of the amount of new scientific knowledge produced by researchers. The impact of this new knowledge can be measured by the number of times publications have been cited by other scientists in subsequent work. Impact, however, cannot be automatically equated with quality. A particular form of estimating the potential quality of scientific papers is to relate this to the prestige and impact levels of the journals in which these are published. These journal impact factors can also be used to compare the citation performance of research groups within specialist fields.

The validity of bibliometric indicators is much greater at the aggregate levels of research groups, university departments and research institutes and should be applied with extreme caution when measuring or comparing the performance of individual scientists. Bibliometric indicators are not intended to replace peer review, but rather to make research visible and debatable, ensuring that experts are sufficiently informed to make sound judgements. Publication-based evaluation, however, considers purely the research aspect of institutional scientific activity and should, therefore, be seen as only a partial indicator of overall scientific performance.

The main constraint to the general validation of bibliometric techniques is the limited availability of databases and other information sources providing reliable and comprehensive raw data for analysis, particularly with regard to research carried out in developing countries. The potential of web-based electronic sources for providing comprehensive and accurate production and citation data for bibliometric analysis coupled with the capacity of the Internet to integrate information from a large number of different sources, promises to revolutionize the way indicators are constructed by eliminating many of the methodological constraints experienced today.

1. Introduction

The worldwide preoccupation with "value for money" in science requiring the rationalization of dwindling support for scientific research, has led to increased use of quantitative data by policy makers. Indicators based on the statistical analysis of quantitative data provided by the scientific and technological literature have been used to measure scientific activity since the beginning of the 20th century. The term "bibliometrics" was first introduced in 1969 as a substitute for statistical bibliography used up until that time to describe the field of study concerned with the application of mathematical models and statistics to research, and quantify the process of written communication. Evaluative bibliometrics is a term coined in the seventies to denote the use of bibliometric techniques, especially publication and citation analysis, in the assessment of scientific activity.

Research evaluation is not the only area of science studies where bibliometrics has a traditional role to play. These techniques are also used extensively for studying the interaction between science and technology, in the mapping of scientific fields, and for tracing the emergence of new disciplines, as well as in the development of foresight indicators for competitive advantage and strategic planning. Bibliometrics is also relevant to other fields. Economists and historians of science, for example, use bibliometric indicators to measure productivity and eminence.

Bibliometric analysis of scientific activity is based on the assumption that carrying out research and communicating the results go hand in hand. Scientific progress is attained by researchers getting together to study specific research topics, steered by the previous work of colleagues. The classic input-output model used to describe the scientific research process suggests that publications can be taken to represent the output of science. Publications, most commonly in the form of the refereed article and the scholarly monograph, are regarded as the definitive statements of the results of research projects. This production can be quantified and analyzed to determine the size and nature of the research carried out. Studies can be performed at macro level to measure global, regional, or national trends or at the micro level of institutions or groups.

Indicators of scientific research can be divided into two main groups: the input indicators such as money spent, equipment used or personnel employed while output indicators such as the literature-based indicators already mentioned, represent the results and outcomes of the research process. Indicators are either absolute or relative. Absolute indicators refer to one particular characteristic of research activity such as number of articles published, number of citations or the amount of money spent while relative indictors show the relationship between two or more aspects such as number of articles per research group or the number of citations per paper. The latter set of indicators is generally more useful in research evaluations due to their ability to establish compound relationships between inputs and outputs such as the amount of money spent per group per article or the productivity of research groups in terms of the number of articles published per group.

Bibliometric indicators are more powerful at higher levels of aggregation and are more suitable for analyzing patterns in a large set (a faculty or large research team) and less suitable for the evaluation of individuals or small research teams. Consequently, the validity of bibliometric indicators when applied to small data sets is questionable making peer review judgements imperative at this level. Whatever their level of aggregation literature based indicators should not be used by non peer policy makers who do not have the necessary background knowledge of the research area or research groups concerned. Interpretation of quantitative data must go hand in hand with qualitative assessment procedures.

At all levels of evaluation no indicator should be taken in isolation. A series of indicators representing the different facets of scientific activity should be employed. When these partial indicators converge to give a unified picture, their validity is strengthened. Some examples of these partial indicators refer both to input into the research process, such as the level of research funding, and also to impacts resulting from the research process. Examples of the latter are non-bibliometric impact indicators such as recognition in the form of prizes or invitations as keynote speakers in major international meetings.

Conceptual and methodological problems associated with finding appropriate output measures arise from the intangible nature of much of the output of basic research activities. Nonetheless, publication and citation data have proved meaningful for measuring scientific output and its impact on the course of scientific research. The number of publications that a research group produces is taken to represent their scientific production and their primary contribution to the generation of new knowledge. Contributions to scientific knowledge take the form of new facts, new hypotheses, new theories or theorems, new explanations or new synthesis of existing facts. The number of times this new information is cited by the authors of later publications measures the impact of their work on the advancement of research in their specialized field, and sometimes, even in other areas of knowledge. It is also indicative of the amount of recognition they enjoy from other members of the scientific community. The reward system theory of science implies that scientists must share their results in order to gain recognition from their peers. Furthermore, the number of publications and citations a research group receives is associated with their visibility as scientists. However, not all published scientific work is equally visible. The level of visibility depends greatly on the place and language of publication as well as the field in question. Work that is not internationally visible will have little chance of being picked up by scientists other than those in close communication with the authors in question. The inclusion of the group's publications in international databases is also a factor affecting their visibility, particularly as these sources are used extensively for the generation of bibliometric indicators.

Over the last decade impact factors (IF) of scientific journals have gained importance in scientific work and information management, as well as in research management and policy. IF is used as an indicator of journal performance and as such has a role to play in the evaluation of research groups, institutes and even countries. Quality journals in science generally contain coherent sets of articles with respect to content as well as professional standards. This coherence stems from the fact that most journals are nowadays specialised in relatively narrow sub-disciplines and their ‘gatekeepers’ (editors and referees) share views on questions like relevance, validity and quality with the invisible college to which they belong.

An important consideration, therefore, in bibliometric studies are the channels used for the dissemination of research work and their coverage in widely accessible bibliographic databases. This latter point is even more important when considering impact indicators due to the fact that only one series of databases, the Citation Indexes produced by the Institute for Scientific Information (ISI) (Philadelphia, USA), is available for citation analysis and for the production of journal impact factors. This service includes only a small proportion of journals published worldwide, restricting its coverage to a few thousand highly cited, mainstream journals.

In the present study we look at one important application of bibliometric indicators, institutional research evaluation based on the analysis of the publication and citation outputs of groups of researchers. The role of journal characteristics, such as the journal impact factor, in literature based evaluations is also described. We concentrate our discussion on the natural and life sciences where bibliometric indicators have reached a higher level of development than in other areas of human knowledge. Special attention is paid to the theoretical foundations of indicator production and the different methodologies available for their construction.

2. Bibliometrics as an evaluation tool

With the advent of "Big Science" bibliometric techniques found a new application in the realms of science administration as a research management and policy tool. Previously, bibliometrics had been the little known domain of librarians, sociologists and historians of science. The need for a relatively quick, easy and inexpensive alternative to peer review for evaluating research performance led to the "discovery" of bibliometrics by science policy specialists and the emergence of a new field of study dedicated to the quantitative study of all aspects of science activity. This new field of scientometrics attracted specialists from different backgrounds, such as mathematicians, information professionals, computer scientists, psychologists, as well as researchers from the natural and medical sciences with a special interest in the study of their own disciplines. The widespread interest in this new field led to the creation in 1978 of its own journal, aptly named *Scientometrics,* and in 1995 to the formation of its own international professional society, the International Society for Scientometrics and Informetrics. Early in the 1960s the introduction of the *Science Citation Index* (SCI) had given bibliometrics a great methodological push. Science indicators research has also been instrumental in the development of the field of scientometrics from the seventies onwards.

Apart from the theoretical and applied research aspects of the field, bibliometrics and scientometrics also give support to countless evaluation exercises performed by tenure, promotion, and awards committees all over the world, as well as by government science policy-makers. While never intended to replace peer review the adjunct of bibliometric indicators make for better-informed expert decisions with respect to budget allocations and in the definition of research agendas and strategic goals. Most bibliometric evaluations of papers, journals and institutions correlate well with peer review appraisals suggesting that bibliometric indicators are generally accordant with the intuitive notions of knowledgeable scientists, as well as with the cognitive state of the art of particular research fields. Nonetheless, rather than bibliometrics being championed as a cheap alternative to peer review, the two methods offering different viewpoints on a common problem, should be considered complimentary and, wherever possible, used concurrently, especially in small scale evaluations.

The expansion of automated bibliographic information services linked to the exponential increase in the volume of scientific literature has presented greater opportunity for the application of bibliometric indicators in research evaluation. This, in turn, has required the design and implementation of better systems design and software development for the handling of large quantities of data and the application of algorithms for the calculation of a wide range of indicators. As these indicators have become more accessible, their weaknesses and strengths have become better understood.

An important and relatively recent application of bibliometrics is in program evaluation. Mapping a field, for instance, before a program is launched, immediately after the end of the program and, perhaps, a few years later furnishes relevant information on many aspects of the field under study, such as the occurrence of cognitive and structural changes. In funding programs too, analysis of scientific publications before and after the funding period can give important insights into its effect on the generation of publishable results.

Although bibliometrics is now a routine tool in evaluations, its use still has its critics. The fact that hard techniques are applied to one important field of human activity, namely the search for new knowledge that are subject to certain social control and coercion, is frequently the basis for censure. Quantitative studies of science then are often reproved for a reputed lack of theoretical foundation. In particular, the absence of a theory of citing is frequently debated, suggesting the need for a more secure epistemological footing to support this practice. Nonetheless, the extensive body of experience gained in the application of bibliometrics in different disciplinary contexts has proved effective for the provision of reliable and useful data for science policy decisions. Interestingly enough, applied techniques, such as the mapping of science, when based on clearly formulated assumptions, have given rise to new theoretical perceptions of the structure and development of science. Useful insights have come also from an increasingly critical user group. Given that the applied side is an important driving force in scientometrics, user feedback has undoubtedly helped to advance the field. For this reason, current research is focussed on the development of new and more powerful literature based indicators required by the user population, as well as on the advancement of fundamental aspects of the field validating it as a *bona fide* research area respected by the broader scientific community.

2.1 Role of bibliometrics in institutional evaluation

In many countries stagnating expenditure on higher education coupled with a growing intake of students in many universities, limit the possibilities for research funding. Furthermore, a growing culture of accountability in research environments is forcing scientists and teachers to become more and more productive. Funds are assigned according to performance. Research evaluation and research excellence are bywords in today's academic climate. Traditionally, assessment of scientific research has been limited to peer review during the grant awarding process or during evaluations for promotion or tenure. Today bibliometric techniques are increasingly used as an intrinsic component of a wide range of evaluation exercises. The present tendency is for institutions to be graded more on the visibility of their products then on their long-term reputation or resources.

The ability of publication and citation analysis to encompass different levels of aggregation makes it a technique ideally suited to national and institutional studies. Nonetheless, literature based indicators are appropriate only for institutional settings that reward publication and only for those activities that produce written knowledge. The fact that the role of written knowledge is influenced by cultural and socio-economic aspects, as well as cognitive determinants that vary between fields of science and between different institutional settings, is considered their main theoretical constraint. Some institutions, for example, recompense behavior that reinforces the reward system of the international scientific community with their own internal reward structures. Others may set their own standards and goals.

Some indicators established globally for the evaluation of scientific performance might not be adequate for a fair or realistic assessment of certain research scenarios. Scientific output indicators based on mainstream publication in international journals should not be taken as the only bibliometric indicator for the evaluation of applied research in developing countries where publication in national journals in the local language is the norm. Disciplinary considerations are paramount. For researchers in the social sciences and humanities, monographs and books are important dissemination channels for research results. Technological research results are published mainly in congress proceedings, reports and patents, and are better represented in this type of gray literature than in mainstream journals. The output of technological and innovation research, in many cases, is not written up as such but appears as designs, applications, models or know-how. In these instances, literature based indicators, clearly, have little meaning.

An important consideration in any exercise of institutional evaluation is that results and recommendations to policy makers should have the general acceptance of the researchers concerned. Consequently, scientists and research managers should be included in the team responsible for the planning, execution and analysis of the research activity. Without the involvement of these key players, the evaluation exercise is unlikely to receive validation by the other members of the research community.

Institutional evaluation should be a continuous process. Ideally, procedures should be in place for the systematic monitoring of research performance and other fundamental scholarly activities. To accomplish this, institutions should develop their own data-system and make it available through the local intranet. In this way information is continually available for consultation by academic staff and other internal users, as well as for providing the raw data for the periodic generation of bibliometric and other scientometric indicators required for evaluation exercises. In practice, most evaluations are focussed on the short-term, often covering only three or four years. This is understandable otherwise results span too long a period for them to be useful for science managers. Nonetheless, their ultimate value can be measured only over the medium and long term.

In institutional evaluation exercises, scientific output and impact are related to input measures, such as research expenditure and the number and categories of academic staff. When carrying out comparative studies, other factors are considered, such as differences in the institutional academic and administrative structures, educational models, etc. Consequently, before deciding upon the procedure for collecting bibliometric data it is necessary to consider the internal institutional research structure. While research administration of many universities follows the traditional departmental structure, the increase in multi and interdisciplinary research, often organized in a program structure, has given rise to research groups formed by members of different departments. Research groups, rather than individual scientists, are today targeted for the allocation of research funds. For this reason, the research group is the most common unit for bibliometric analysis in institutional evaluations. This in turn has produced a wave of interest in scientometric research focussed on the identification of research groups by coauthor analysis and its corroboration by expert opinion. Notwithstanding, the research performance of any aggregate of scientists can be assessed using bibliometrics. This aggregate is often termed a "unit" which can be taken to represent any given set or sets of scientists depending upon the objective of the evaluation.

While no absolute quantification of research performance is possible, valid and useful comparisons can be made between research groups working in the same fields. When making comparisons between groups it is essential to apply indicators to matched groups, comparing like with like, as far as possible, and to give careful thought to what the various indicators are actually measuring. It is also important to study not only the similarities between groups but also the differences, especially those that could be directly influencing the research performance.

2.2 Methodological considerations

Different data sources are available for bibliometric analysis. Before the advent of bibliographic databases quantitative analysis of the scientific literature was done manually directly from the published material. This is still possible in small-scale studies but it is labor-intensive and a lot of time is spent entering and checking data. In institutional evaluations it is possible to use internal documents such as research reports or annual reports which are often available through the institutional libraries. Alternatively, where researchers are expected to produce periodic progress reports on their research work as part of an institutional or program requirement, these can be used for analysis. When information is required in addition to that contained in the records of bibliographic databases, such as funding sources or acknowledgements, the original publications have to be located and consulted.

Large-scale analysis, on the other hand, requires the use of commercial, public or institutional bibliographic databases either online through the Internet or using a database vendor service. The most viable offline option involves consulting the CD-ROM versions. Online access requires a fast option for downloading records unless offline downloading and subsequent delivery by email is an option. The mainframe databases of the producers contain the most up-to-date information that will not typically be the case with the CD-ROMs. Special software is now available for the downloading and handling of bibliographic records, in some cases *ex professo* for bibliometric purposes.

Databases providing numerical data for indicator production, the so-called science indicators databases, are also available for purchase usually corresponding to records with specific characteristics retrieved from the mainframe database. Many of these can be customized with respect to the characteristics of the records required (all records from a particular institution or for a particular group of scientists). Also special provision can be made by the database producer or vendor for certain analytical flexibility that is not possible using records recovered from the online version. Likewise retrieval software in the CD-ROM version generally places certain limitations on the types of data that can be recovered and there are fewer options for data manipulation. Many databases offer abstracts in addition to bibliographic data that can then be searched for more specific information regarding research content. Some even offer full text articles.

A third option is to contract the services of a handful of agencies throughout the world that specialize in gathering and organizing literature–based indicators for evaluation purposes such as the Institute for Scientific Information (ISI) in Philadelphia; CHI Research, Inc. in New Jersey; the Information and Scientometrics Research Unit (ISSRU) in Budapest; and the Center for Science and Technology Studies (CWTS) at Leiden University in the Netherlands. These agencies also carry out innovative research on the development of new indicators and indicator data sources. Little standardization exists with regard to precise methodologies and sources used for indicator generation between the different producers making comparisons generally invalid. Although most indicators are generated using records provided by ISI, the precise data sources employed are often based on different journal sets. Variations also occur in specific datafields, particularly in the address field and types of documents.

Although created initially as a new information retrieval tool based on citations, the particular attributes of the SCI have made this the most frequently utilized bibliographic file for bibliometric analysis. SCI is a multidisciplinary database indexing cover to cover approximately 3,500 international journals, in most subfields of the natural and life sciences. This is in contrast to most other international databases, such as Chemical Abstracts, BIOSIS, or Medline that cover specific disciplines. SCI includes all author addresses while the majority of the best known international databases include the addresses of only the first authors of papers. This is an important consideration in collaboration studies and when comparing the scientific production of different countries and regions. However, it is the inclusion of all papers cited in each of the articles indexed in the database that makes the SCI unique.

ISI also produces citation indexes in the social sciences, *Social Sciences Citation Index*, and in the arts and humanities, *Arts and Humanities Citation Index*. In these fields, however, the applicability of citations to assess research impact is questionable for which reason these indexes are not used as extensively as in the natural sciences (see *Citation Measurements*). In addition, the fact that SCI processes few books, proceedings and other types of documents limits its usefulness in disciplines where the journal article is not the main vehicle for communication of results. The disciplinary orientated international databases tend to cover a wider range of topics within the discipline, of document types and process a far greater number of national journals. This gives them a broader information base for the bibliometric analysis of research performance in all fields and in countries whose journals are poorly represented in the SCI, such as those from the developing world.

In general, bibliographic databases designed not for bibliometric purposes, but rather to provide users with a way of accessing and controlling the abundant literature available in their specialized fields, are by no means ideal sources of data for analysis of research performance. Lack of standardization in searchable fields and the inability to recover information on important aspects of the published documents for bibliometric analysis are just two of the limitations. Such is the case of author affiliations, so important in institutional analysis. Assigning authors to their institutions is somewhat problematic as addresses are often incomplete and sometimes incorrect. There is a special problem with the names of institutions from non-English speaking countries as most of the big international databases are developed in English. Furthermore, articles do not necessarily have a specific field assignation. This difficulty has been resolved to some extent by assigning articles to the disciplinary areas of the journals in which they are published. Special classifications of journals have been developed for this purpose by organizations such as CHI Research. ISI and others have developed over the years novel indicators to enrich the battery available for use in evaluation studies. A well known example is ISI's journal impact factor (IF) which determines the impact of journals based on the number of citations they receive in relation to the number of articles published over a specific time period (see *Journal Impact Factors*).

To achieve a higher degree of standardization in bibliometrics is by no means an easy task. Responsibility rests not only with bibliometricians and database producers, but also with journal publishers and editors who need to formalize and standardize their procedures and methods more rigorously. While some journals link individual authors with their institutions, others list all institutional addresses without making explicit the relationship between the authors and the addresses. In many journals it is difficult to identify different document types and some have different definitions of a particular document type. Letters are a case in point.

The bibliometrician, then, must know fully the data sources available and carefully select those most likely to answer the questions under study. For instance, it is well known among specialists that different results will be forthcoming when consulting the SCI in CD-ROM version compared to doing a search online via the SCISEARCH database. Adequate strategies then have to be developed for the downloading of the required records and their subsequent manipulation. Some aspects of each record that may not be included in an independent field in the original database, such as the countries that normally form part of the author address, need to be separated out and individually coded. Data need to be validated as far as possible and adjustments made where necessary. This often implies checking with the original journals, when practically feasible, or with reference works. The advantage of using the few "bibliometric" databases available for consultation is that this alleviates the problem of extensive cleaning up of data, classifying and coding fields necessary before initiating the process of bibliometric analysis.

The best strategy for collecting a complete and accurate publication list for the unit to be evaluated is, perhaps, to use a combination of data sources. The use of publication lists provided by the members of the unit, together with references to documents retrieved from a suitable bibliographic database by searching on the names of the group members, should cover most of their scientific output. Annual reports are often incomplete and should therefore be used only as complementary sources of publication data. Citation data should be collected for all published material and not restricted to publications in journals covered by SCI. Work published in journals and series not covered by SCI can well be cited by SCI covered sources.

The next decision to be made is what type of output and impact indicators should be applied to the unit under study. This will clearly depend upon the precise objectives of the evaluation as well as the research fields concerned. Large differences in applicability of bibliometric data have been found even within fields in the natural and life sciences. The particular characteristics of the unit to be evaluated such as the relative prominence given to teaching and research activities, also have to be taken into consideration especially in comparative studies. New indicators are continually being devised, some of which may only be relevant in particular institutional settings. Below we describe only those that are well established and which are generally applicable to the majority of research scenarios. In most instances a combination of production and citation indicators is employed.

3. Output evaluation

The number (quantity), type and distribution of publications are the most commonly applied bibliometric indicators of scientific output. The production of a research unit is its number of publications. Productivity is expressed as the number of publications per person-year equivalents for research.

The production of research publications is highly skewed. In any given field of research, a small number of high producers will be responsible for a significant percentage of all publications in the field. The distribution of individual productivity was first examined by Lotka whose power law [equation (1)] is one of the major regularities studied in bibliometrics:

$$x^n y = c \quad (1)$$

where y is the proportion of authors making x contributions each; n and c are parameters depending upon the time span covered by the bibliography, and probably the field being analyzed.

Important decisions have to be taken before embarking on publication analysis, the most fundamental of which refers to the types of publication to count and their relative worth. Other important considerations are what time spans to use and how best to assign credit to individual authors or institutions in multi-authored papers or books.

3.1 Weighted values of publications

Only appropriate kinds of publication should be counted depending upon the prevalent communication channels of the fields under study. While many researchers in the humanities prefer to publish results in book form rather than as articles, in the applied sciences, such as engineering research reports and patents are more common forms of output. Where, for example, counts of articles and books are mixed, the values assigned to books need to be weighted. Review articles also need special consideration. Output comparisons between scientific and medical fields, on the one hand, and the social sciences and humanities on the other, depend to a large extent on the relative weightings assigned to the different types of publications used by researchers in these two distinct areas of knowledge. One book has been deemed equivalent to anywhere between two to six articles.

Even in the experimental sciences where refereed articles are the obvious output medium, the relative importance for the evaluation exercise of other document types need to be assessed. The SCI database classifies journal papers into several types of which usually only three or four are considered as significant contributions to research. Articles, notes, reviews and letters are commonly included in both scientific output and citation counts. Letters are sometimes excluded depending upon their function in the particular field under study. Other publication types such as meeting abstracts, discussions and editorials are useful for bibliographic retrieval but are generally excluded from bibliometric analysis.

Distinctions are often made in evaluation exercises between papers published in SCI journals and those published in non-SCI journals. The first category refers to highly cited, highly visible journals for which citation data are available. Publication in the second category of journals is unlikely to produce the same level of citation as in the more prestigious SCI journals due to reduced visibility within the scientific community. The fact that authors compete for publication in SCI journals bestows on these a certain intrinsic element of quality regardless of their citation levels.

Even when considering the output of a single author or group of authors in a given field each publication may not represent an equal contribution to science. Not all articles and more especially not all books, contain original research. Some are aimed at discussing current developments or matters of professional concern. Books in the humanities are more likely to represent original research contributions than those in the sciences where textbook writing is more prevalent. For this reason, some solid conceptual basis is needed to decide what should be counted as a publication. Furthermore, the amount of work contained in a single article or book varies not only between fields but also by individual choice. Consequently publications are sometimes weighted according to the number of published pages, as well as in relation to the number of authors.

The increasing trend for research performance to be measured principally on production output and citation impact has produced a phenomenon known as salami slicing, whereby some scientists spread their results over several less notable papers instead of concentrating these into one more weighty and significant article. The importance of peer review in these cases is paramount. When the unit to be evaluated consists of a large number of scientists with a significant number of papers, the "salami slicing" effect will cancel itself out unless of course this practice is widespread.

3.2 Time span

A fixed time span of one year or more is appropriate for the analysis of the publication output of research units such as groups or institutions. Four-year periods are often used (see *Role of bibliometrics in institutional evaluation).* An important consideration when comparing publication histories of groups of scientists, is that their professional ages (normally taken as the number of years since being awarded their PhD degrees) must be similar. In comparing the publication outputs of research groups, significant differences in the years lapsed since the groups were formed should be taken into consideration.

Time series analysis can be applied to supplement average values calculated for the entire time span. Indicators are studied on a year-to-year basis in order to trace the evolution of a group's production (and impact). Examples of these are total number of publications published annually; number of SCI publications published annually; percentage of SCI publications published annually.

Trend analysis is also useful for determining the effect on scientific output of variations in the research environment, such as changes in research budgets and in institutional goals, and the implementation of reward and incentive schemes. Publication output can vary considerably with time also for individual reasons. This appears to be true for both high and low producers.

3.3 Assigning publications to different research units

One of the frequent aims of institutional research evaluation is to analyze the contributions made by particular scientists or groups of scientists. In today's climate of increasing collaboration in science the evaluator has to decide the most equitable way of assigning credit to each and everyone of the authors signing a paper. This situation arises when a research unit publishes articles in collaboration with scientists from other groups both from within their institution and from outside.

Several procedures have been put forward, two of which are the more frequently applied. Co-authored publications are counted either as one publication for each of its co-authors (total author counting) or as fractional counts (fractional counting) assigned to each author based on the number of co-authors. When a paper has A authors, for instance, each author is attributed with one point. Alternatively, each of the A authors receives a score equal to $1/A$. Another method involves giving credit to only the first author of the paper (first author counting). The popularity of this method is due to the fact that in most formats ISI citation data (not publication data, however) are based on the first author only.

Table 1 gives an example of a small set of four articles, and their respective authors.

Table 1

	Authors		
Article 1	a_1	a_2	a_3
Article 2	a_4	a_2	a_3
Article 3	a_3	a_4	
Article 4	a_5	a_4	

First author counting (FA), total counting (TO), and fractional counting (FR) lead to the following, different, rankings between authors:

FA: $a_1 = a_3 = a_4 = a_5 = 1$; while $a_2 = 0$
TO: $a_3 = a_4 = 3$; $a_2 = 2$: $a_1 = a_5 = 1$ (a_2 moves from last to third)
FR: $a_4 = 8/6$; $a_3 = 7/6$; $a_2 = 4/6$; $a_5 = 3/6$; $a_1 = 2/6$

In this example fractional counting refines the ranking obtained by total counting, while first-author counting does not give an accurate picture of the relative contribution of the authors.

Which of these methods to use is an important decision as different accrediting procedures may produce significantly different results. Help may be available in the form of publication policies of a particular group, department or institution. In some academic environments, for example, it is customary for the names of senior faculty to appear at the end of the author list, more in virtue of their position than any contribution they have made to the study. The order in which authors' names appear in a paper is often field dependent. In some fields alphabetical listings is preferred while in others the first author is the scientist making the greatest contribution to the research. When deciding upon a particular procedure these practices should be taken into consideration. In the absence of these, different counting methods should be applied and compared. Although the simplest method to apply, first-author counting is not recommended. It is particularly unfair when authors are systematically ordered alphabetically, or practice ‘noblesse oblige’ (putting the senior author last).

4. Citation Measurements

The aim of citation analysis is generally held to be that of injecting a "quality factor" into the evaluation of scientific performance. The level of interest that colleagues show in the published research of a group of scientists is seen as one way of estimating quality. However distinction should be made between quality, importance and impact. While the first is an inherent property of the work, the last two are based on external appraisals. Importance refers to the potential influence on surrounding research activities while impact reflects the actual influence. Judgment of the basic quality of scientific research can be made only by peers. To operationalize the general concept of quality in scientific research is virtually impossible as this may refer to a variety of values - cognitive, methodological and esthetic. But even if citations do not provide an indicator of the quality of scientific papers, they do reflect, at least partially, the impact they have on the scientific community. "Imperfections" in the scientific communication system result in the importance of a paper not necessarily being identical with its impact.

The most important measure of impact is the number of citations, or, on an aggregate level, the average number of citations per paper. Short-term impact is indicated by counting the number of citations received by the publication during the first few years following its appearance, for example during the first three years or during the third year after publication. Three-year or four-year citation counting periods are often used. The period over which citations are counted is termed the citation window. Long-term impact considers a longer time span.

There is a time lag between work being published and its coming to the attention of researchers in the field who may wish to cite it. Obviously the time that elapses between work being completed and its being cited will be partially determined by the speed and place of publication. Although scientists are naturally interested in their results being published expeditiously rapid publication is usually regarded as less important than the prestige or readership of the journal selected for publication. Interest in speed of publication varies with discipline. In areas where research is developing rapidly how quickly results can be communicated to the specialist community is of the utmost importance for laying claim to results before other groups do. Rapid publication is also a requisite for the prompt citation of research work.

Age distribution of citations within a specific research area will determine the speed with which publications are normally cited. There are substantial differences in the year when maximum citations are achieved, both between fields of research and within fields. Some papers make a quick but fleeting impact on their fields, others have a slower but more durable impact. Important paradigm defying research will continue to be cited until it is superseded or becomes part of the established body of knowledge in the field. Short-term impact is related to the visibility of research groups at the research front and can be ranked with other visibility indicators, such as international contacts, awards, invitations to participate in prestigious international meetings, among others. Short-term impact does not necessarily guarantee long-term impact as many attractive theories fail to withstand the test of time.

Long-term impact is not a very useful indicator in the evaluation of research groups, partly because certain of these may have ceased to exist or have changed their research interests. Furthermore there is no guarantee that groups will perform equally well (or badly) over prolonged periods. Research groups cannot be required to carry out work with long-term impact but they can be expected to take part in scientific discussions in the field. For this reason, short-term impact is of primary importance in institutional evaluation.

Following the argument that it is not the great mass of small impact papers that contribute most to scientific progress, but rather those few key papers having a great impact, the presence of highly cited papers is an important performance indicator. This can be expressed as the annual percentage of highly cited papers (those perhaps receiving more than a certain number of citations per year) and is useful when comparing the citation performance of matched research groups or institutions. Nonetheless, this indicator is seldom calculated in purely bibliometric analysis as the presence of highly cited papers is an attribute that peer review will highlight.

Most authors include citations to their own work in their papers. This is to be expected as scientists build on their previous work when designing experiments making internal citations essential, unavoidable and justifiable. Self-citations account for some 10% of all citations. Although citations by one of the co-authors of a paper does not constitute impact on peers, these are rarely eliminated from citation counts. The final estimation of the impact performance of a research group, for instance, will be the result of a comparison with that of the field in general. This is normally done using ISI data on field and journal impact factors (see *Journal Impact Factors*) and citation scores which include self-citations. For most evaluation exercises the elimination of self-citations from ISI data is not a practical option.

The variability to be expected in citation counts is enormous. Review publications and work on innovative techniques and methods are usually highly cited. The absolute number of citations depends greatly on the particular characteristics of the research topic or field in question. Scientists publishing in fields with a large volume of research output, such as molecular biology, are in a better position to be cited than those working in disciplines such as mathematics, where a research paper is longer in the making and is likely to contain only a few references to previous work. These differences will be reflected not only in the number of citations per paper but also in the impact factors of journals in these different fields (see *Journal Impact Factors*).

Objections raised regarding the use of citation data for evaluation purposes have focused on assumptions regarding the citation process. Many random factors unrelated to the scientific importance of research contribute to the final outcome of papers being cited, such as certain references being close at hand or the choice made between equally relevant papers. Thus when small numbers of papers are involved, chance factors may obscure a real difference in impact. As the number of papers analyzed increases, the relative contribution of chance factors decreases and that of real differences increases. Consequently, studies involving larger sets of papers give more reliable results.

As with production data citation analysis is considered a fair evaluation tool for those scientific sub-fields where publication in the journal literature is the main vehicle of communication but only when the limitations of the databases used for the evaluations are fully understood and taken into account for the final conclusions. Furthermore there is powerful evidence to show that citation measures correlate well with faculty performance ratings.

Citation analysis is subject to a certain level of inaccuracy. Errors and variations occur in the details of the cited articles in the SCI many of which can be related to inaccuracies in the original publications. The presence of a specific mistake in a citation to a particular article occurring in more than one citing publication suggests that authors may copy references directly from lists in other papers thus propagating errors. Other discrepancies appear to be due to irregularities in the internal standardization process at ISI or are simply keyboarding errors.

5. Journal Impact Factors

The study of the use and relative impact of scientific journals is one of the important applications of citation analysis. Research on citations was given a significant impetus by the publication of ISI's Journal Citation Reports starting in 1976. These reports, currently available on the web, in CD-ROM and on microfiche, provide a wealth of indicators and statistical data on the citation patterns of journals covered by ISI's series of citation indexes. Frequently applied in research evaluation exercises are the international journal lists grouped by discipline and ranked according to impact factor. The ISI impact factor (IF) is basically a ratio between citations and citable items published in journals. It measures the frequency with which the average cited article in a journal has been cited in a particular year.

The ISI impact factor [equation (2)] of a journal in the year Y is calculated as:

$$\frac{CIT(Y,Y-1)+CIT(Y,Y-2)}{PUB(Y-1)+PUB(Y-2)} \qquad (2)$$

where the number of citations in year Y to papers published in the year Z is denoted as CIT(Y,Z), the number of publications in the year Y by PUB (Y).

In order to calculate an analogue of the ISI impact factor for a non-ISI journal, this journal is simply added to the pool of ISI source journals. How often this particular journal is cited by ISI journals (during the period under investigation) is calculated and added to the number of times the journal cites itself. This is then divided by the number of articles published by the non-ISI journal. Although this is a simple procedure there are two caveats. Firstly, ISI always includes journal self-citations, so this must be done here too. Secondly, if this new impact factor is used to compare this non-SCI journal with ISI-journals, the ISI-journals' impact factor must be recomputed, because the pool of journals has changed.

Selection of appropriate citation and publication windows is crucial for calculating impact factors. Mean citedness, i.e. impact factors, can be calculated using either a synchronous or a diachronous approach, and with different time windows for publication and citation data. The ISI impact factor is synchronous involving a single citation year and two publication years. The term 'synchronous' refers to the fact that the citations are all "given" in the same year. These impact factors "mix" different publication years rendering them more robust when the object under study is the journal itself. Synchronous indicators better represent the permanent impact of journals whereas diachronous indicators characterize their actual impact. Consequently when the article (or the scientist who wrote it) is being evaluated diachronous impact factors are more commonly employed. Furthermore they can also be calculated for one-off publications such as books containing contributions by different authors and conferences proceedings.

The following formula is used to calculate the n-year synchronous impact factor of a journal J in the year Y:

$$IF_n(Y)=\frac{\sum_{i=1}^{n} CIT(Y,Y-i)}{\sum_{i=1}^{n} PUB(Y-i)}$$

The n-year diachronous impact factor of this journal for the year Y is:

$$IMP_n(Y)=\frac{\sum_{i=k}^{k+n-1} CIT(Y+i,Y)}{PUB(Y)}$$

where k = 0 or 1, depending on whether the year of publication is included or not.

Many specialists warn against the use of journal impact factors as surrogates for actual citation performance. One of the main reasons put forward is that large differences exist in quality and citation scores between papers published in the same research journal. The presence of highly cited papers tends to artificially bump up impact factors. Furthermore, all citations are treated as equal regardless of the citing journal. Some fields that are useful for science do not receive many citations. Consequently, if the impact factors and other citation-based measures were to become the main criteria for journal quality then whole sub-fields would be eliminated and others would have their worth undermined.

As is true for the number of citations received by an individual paper the IF of journals is similarly affected by factors unrelated to the scientific quality of the articles. Such is the case of foreign journals handicapped for their inclusion in the SCI by a lack of English content. A high quality paper in a less visible foreign journal is likely to receive fewer citations than a poorer quality paper published in a highly visible SCI journal. Consequently, impact factors measure only the (international) use of journals at the research front of their respective fields. Evaluators should bear this fact very much in mind when applying IFs in assessment exercises especially at low levels of aggregation.

6. Relative Impact Indicators

A time series analysis alone is not sufficient to obtain a complete picture of the impact of research groups. In the absence of any absolute standards their performance needs to be compared to that of other research groups in the field. One way of doing this is by calculating the ratio of the average of the group's citations (per article) with the average of the journals in which they have published. Alternatively the ratio of the average of the group's citations with the average of the field (or fields) in which they are active can be determined.

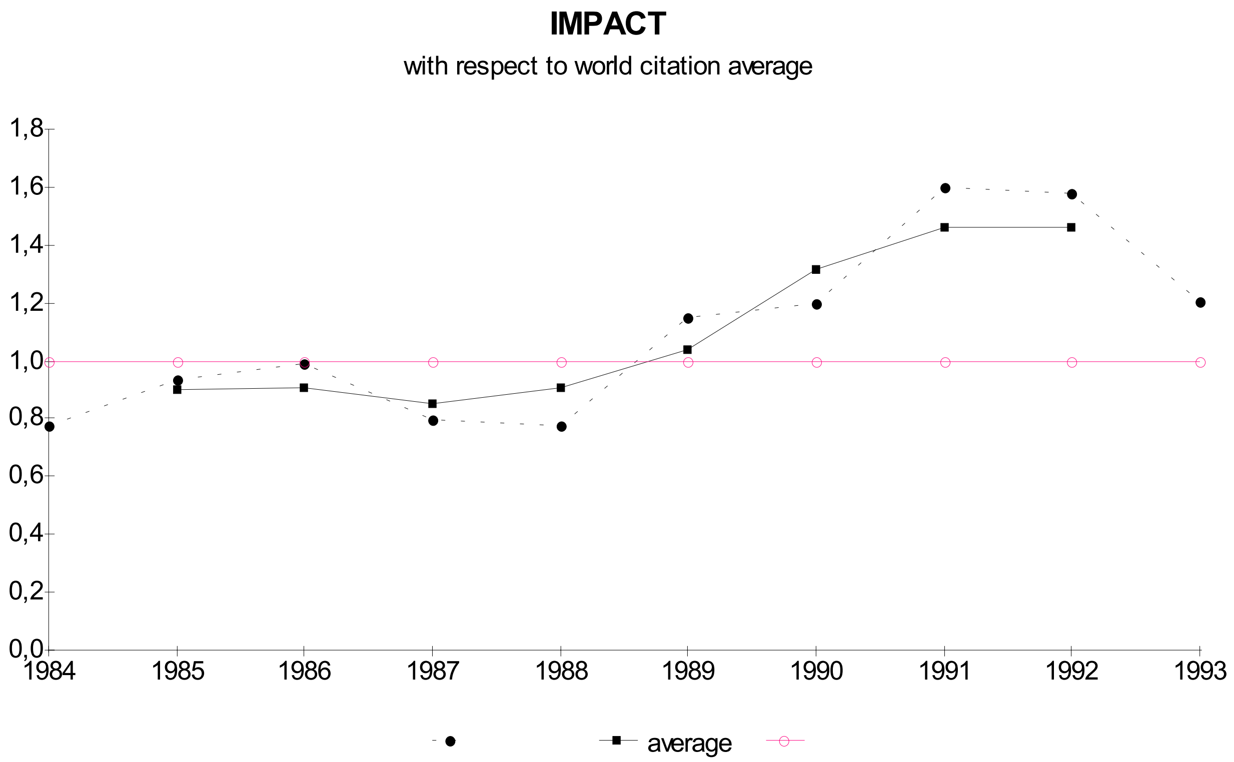
A group publishes articles in a number of journals: the so-called journal package of the group. It is possible to obtain an IF for each journal of this package, i.e. an average number of times an article published in this journal is cited. The impact factor is denoted as **JCS** (Journal Citation Score). The **JCS** or expected impact of a group of articles is calculated using publication and citation data from the JCR. Weighted average IF for the journal package in which the group publishes can be constructed using impact factors of journals (symbol: **JCSm**).

However this does not answer the question of whether or not a group is more cited than similar groups nor whether the group's journal citation score is high or low when viewed from an international perspective. To do this we need to calculate, per sub-field, the average number of citations an article from that field receives, called the world average per field, and denoted as **FCS** (Field Citation Score). Usually a group is active in several sub-fields, and consequently a weighted average is used (symbol: **FCSm**), similar to the calculation of **JCSm.**

The indicator **JCSm/FCSm** compares the weighted-average-impact factor of the journal package with the world citation average of the fields in which the group is active. It corresponds to the ratio of the average impact factor of the journal package to the average world impact of the fields in which the group is active. When the **JCSm/FCSm** is larger than one, the group publishes in journals having an IF that is larger than the world average in the field. In this way we can compare the actual number of citations per publication of a research group with the expected number based on the average JCS values to determine whether this actual number of citations is high or low. The validity of the indicator is based on the assumption that the set of journals used in the evaluation yields an adequate representation of the discipline.

Figure 1 is an example of the impact of the research publications of a Belgian university using this indicator which compares citation scores with the field. Since groups publish in different fields, the figure contains weighted scores. A three-year average line is used to smooth out irregularities and make the general trend more obvious.

Figure 1



When calculating the impact of a field, two approaches are possible: either the average of the impact factors of all journals in the field is calculated (this is called the average impact of this set of journals), or, alternatively, a global average. The latter is the better approach. The difference between these two approaches is shown as follows. If C_i denotes the number of citations (over a certain period) of journal i, and if P_i denotes the number of publications in journal i, then I_i will denote the impact of journal I (citations per publication). The average impact factor is then defined as:

$$AIF = \frac{1}{n}\sum_{i=1}^{n}\frac{C_i}{P_i} = \frac{1}{n}\sum_{i=1}^{n} I_i$$

The global impact, on the other hand, is calculated as:

$$GIF = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} P_i} = \frac{\mu_C}{\mu_P}$$

where μ_C and μ_P denote the mean number of citations, and the mean number of publications.

7. Future trends and perspectives

The development of bibliometric and scientometric techniques has been strongly influenced by the availability of adequate data, as well as by the facilities available for computing large quantities of information. Consequently the transition from print format to electronic media in scientific communication, will also bring about significant changes to the way production and citation indicators are constructed. The Web as a place of integrated scientific and technical communication promises to become not only a major new and improved source of production and citation data but also a legitimate one. For this reason, any advanced bibliometric data-system for institutional monitoring must be sufficiently flexible to take into account rapid developments in electronic publishing and in relevant Internet facilities

Major changes to bibliographic and bibliometric methodologies will emerge with the increased presence of academic journals and other scholarly works on Web. The seemingly limitless capability of the Web for storing and integrating information will make available unified (and hopefully, uniform) data sources for bibliometric analysis, accessible from any PC anywhere in the world. A case in point would be the emergence of local and regional citations indexes and their integration into a virtual world citation atlas. The presence of every scholarly work ever written linked to every work it cites or is cited by in a universal Web based bibliographic and citation database would solve many of the problems plaguing the construction of output and citation measures in today's non-electronic environment.

New indicators will emerge such as those already proposed to legitimize the importance of intellectual input of colleagues recognized in the acknowledgements section of research papers and reports. The "seamless" all-embracing Web environment could lead to a better assessment of multidisciplinary work as well as the identification of new and significant relations between different disciplines.

The problem of the under-representation in international scientific databases of studies from developing and other countries written in non-English languages could eventually become a thing of the (non-electronic) past. The prospect of on the spot translation of non-English articles on the Web, for example, would lead to an increase in the visibility and citation levels of non-English speaking scientists. Online publishing ventures such as the establishment of global repositories for research would also increase the holistic presence of studies from developing and other peripheral countries.

The fact that the Internet accentuates the value of individual pieces of information puts increasing emphasis on the individual article and rather less on the journal, a trend that could diminish the value given to journal impact factors in the short-term and likely to cause their demise in the long-term. Clustering articles on the Web by subject would allow these to be ranked according to their importance for different fields, a process which could identify citations made to inappropriate references as a result of scientists playing the "citation game" of preferentially citing their own or their colleague's work. Furthermore, with escalating networking in science, scientists will become increasingly aware of what their peers are doing. This could bring about a possible increase in the speed with which results are incorporated into the work of others thus reducing citation lags.

We can also look forward to the opportunity for better accuracy in bibliometric data with the automatic softbot checking of bibliographic elements such as titles, authors, directly from their original Web sources. We can also hope for greater access to the results of research from the social sciences and the humanities due to increased international presence through web publishing.

In anticipation of these and other major paradigmatic shifts in scientific communication, a new field of study is already emerging. Cybermetrics or webometrics is the name given to the innovative metric study of electronic communication. At this transitional stage there are numerous discussions on the validity of adapting traditional methods for the construction of production and citation data using electronic formats. One particular concern is the definition of what constitutes a valid publication in Internet. Another issue is how to calculate Web impact factors and their validity in terms of measuring the impact of a particular web space.

But one aspect that will change little in the face of the electronic revolution is the human element. Our poor understanding of the psychological and social processes underlying peer review as well as those involved in the publication and the citation processes, will not be automatically improved by wider access to output and impact data. Notwithstanding we will be able to provide peers with more comprehensive and more reliable bibliometric data to guide and support their decisions and to better defend these before different non-scientific sectors of the community, such as science managers, politicians and the general public.

It should be kept in mind that research evaluation is not an end in itself. It is only an aid to the real goal that is providing those people and institutions with the talent and motivations to carry out scientific research, with the best conditions possible under which to do so. Budgetary and other kinds of constraints make evaluations necessary for the equitable distribution of resources. The evaluation of short-term strategic research as well as the long-term curiosity-driven search for new knowledge demands the same accountability and rigorous standards as scientific research requires of itself. For this reason the challenge is not only for the application of bibliometric and scientometric techniques in research evaluation to keep up with the rapid changes occurring in scientific communications patterns and practices but also to constantly improve the theoretical foundation for the construction of output and impact indicators as an adjunct for peer review.

Bibliography

Cronin B. (1984). *The citation process. The role and significance of citations in scientific communication.* London: Taylor Graham. [An essay, still valid today, debating the meaning of citing practice focussing on science as a social process.]

De Bruin, R.E., Kint, A., Luwel, M. and Moed, H.F. (1993). A study of research evaluation and planning:the University of Ghent. *Research Evaluation* 3(1), 25-41. [Provides a detailed description of how to perform a systematic evaluation of research performance.]

Egghe, L. and Rousseau, R. (1990). *Introduction to informetrics. Quantitative methods in library, documentation and information science*. Amsterdam: Elsevier. [Part III of this book describes the basic ISI citation tools and provides definitions of citation and citation-related concepts.]

Garfield E. (1979). *Citation indexing-its theory and application in science, technology and humanities.* New York: Wiley. [Seminal work by the "Father" of citation indexing.]

Martin, B.R. and Irvine, J. (1983). Assessing basic research. Some partial indicators of scientific progress in radio astronomy. *Research Policy* 12, 61-90. [Convergent partial indicators provide a reasonable estimate of a group’s relative contribution to scientific progress.]

Moed, H.F, Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy* 14,131-149. [The first complete description of the evaluation of university research groups indicating the potential of bibliometric indicators as tools for university research policy.]

Proceedings of the Biennial Conferences of the International Society for Scientometrics and Informetrics. 1987-1999. [Proceedings series reporting original studies in the field of Bibliometrics, Scientometrics and Informetrics many of which are related to indicator production.]

Van Raan A.F.J. (ed) (1988) *Handbook of quantitative studies of science and technology.* Amsterdam: Elsevier. [The main purpose of this compilation is to present a wide range of topics in the domain of quantitative studies, incorporating theory, methods and applications.]

Van Raan A.F.J. (1997) Scientometrics: State-of-the-art. *Scientometrics* 38(1), 205-218. [This review emphasizes the duality of scientometrics as both a basic and applied field, in which the applied side is considered the driving force.]