

The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index

Peder Olesen Larsen · Markus von Ins

Received: 3 September 2009 / Published online: 10 March 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The growth rate of scientific publication has been studied from 1907 to 2007 using available data from a number of literature databases, including Science Citation Index (SCI) and Social Sciences Citation Index (SSCI). Traditional scientific publishing, that is publication in peer-reviewed journals, is still increasing although there are big differences between fields. There are no indications that the growth rate has decreased in the last 50 years. At the same time publication using new channels, for example conference proceedings, open archives and home pages, is growing fast. The growth rate for SCI up to 2007 is smaller than for comparable databases. This means that SCI was covering a decreasing part of the traditional scientific literature. There are also clear indications that the coverage by SCI is especially low in some of the scientific areas with the highest growth rate, including computer science and engineering sciences. The role of conference proceedings, open access archives and publications published on the net is increasing, especially in scientific fields with high growth rates, but this has only partially been reflected in the databases. The new publication channels challenge the use of the big databases in measurements of scientific productivity or output and of the growth rate of science. Because of the declining coverage and this challenge it is problematic that SCI has been used and is used as the dominant source for science indicators based on publication and citation numbers. The limited data available for social sciences show that the growth rate in SSCI was remarkably low and indicate that the coverage by SSCI was declining over time. National Science Indicators from Thomson Reuters is based solely on SCI, SSCI

A preliminary version was presented at the 12th International Conference on Scientometrics and Informetrics 2009 (Larsen and von Ins 2009).

The author sequence is alphabetic and does not reflect relative contributions to the work.

P. O. Larsen (✉)
Marievej 10A, 2, 2900 Hellerup, Denmark
e-mail: pol@webspeed.dk

M. von Ins
Institute for Research Information and Quality Assurance iFQ,
Godesberger Allee 90, 53175 Bonn, Germany

and Arts and Humanities Citation Index (AHCI). Therefore the declining coverage of the citation databases problematizes the use of this source.

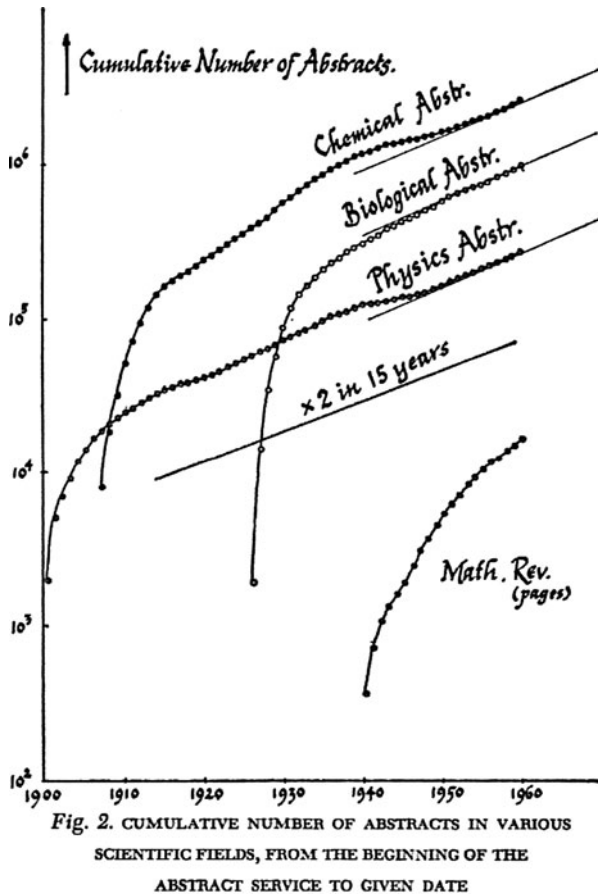
Keywords Growth rate for science · Growth rate for scientific publication · Databases for scientific publications · Coverage of databases · Coverage of science citation index · Coverage of conference proceedings · Number of scientific journals · Little Science, Big Science · Exponential growth · Doubling time · Cumulative values

Introduction

In 1961 Derek J. de Solla Price published the first quantitative data about the growth of science, covering the period from about 1650 to 1950. The first data used were the numbers of scientific journals. The data indicated a growth rate of about 5.6% per year and a doubling time of 13 years. The number of journals recorded for 1950 was about 60,000 and the forecast for year 2000 was about 1,000,000 (Price 1961). Price used the numbers of all scientific journals which had been in existence in the period covered, not only the journals still being published. However, this is not a major source of error. In 1963 Price continued the work using the number of records in abstract compendia for the period from 1907 to 1960. Figure 1 is a copy of the classical figure from *Little Science, Big Science*, with the data for Chemical Abstracts, Biological Abstracts, Physics Abstracts and the Mathematical Review.

From the data Price deduced a doubling time of 15 years (corresponding to an annual growth rate of 4.7%). Price underlined the obvious fact that this growth rate sooner or later would decline although until then there were no indications of this. Price conjectured “that at some time, undetermined as yet but probably during the 1940s or 1950s, we passed through the midperiod in general growth of science’s body politic” and that although “It is far too approximate to indicate when and in what circumstances saturation will begin ... We now maintain that it may already have arrived” (Price 1963, p. 31). Price also discussed the increasing role of the newcomers in science, first of all The Soviet Union and China. He suggested that the doubling time in The Soviet Union for science might be as low as 7 years and that “one may expect it [China] to reach parity within the next decade or two” and that “the Chinese scientific population is doubling about every three years” (Price 1963, p. 101). Subsequently Price stated: “all crude measures, however arrived at, show to a first approximation that science increases exponentially, at a compound interest of about 7% per annum, thus doubling in size every 10–15 years, growing by a factor of 10 every half century, and by something like a factor of a million in the 300 years which separate us from the seventeenth-century invention of the scientific paper when the process began” (Price 1965). However, a growth rate of 7% per year corresponds to a doubling time of 10 years, growth by a factor of 32 in 50 years and of one billion in 300 years, obviously too high.

Price’s quantitative measurements were not completely correct but his investigations were pioneering. As a result of his work Research and Development (R&D) statistics and science indicators have become necessary and important tools in the science of science, research policy and research administration. Publication numbers have been used as measures of the output of research, especially academic research and university research. The basis for the measurement of publication numbers are the big databases for scientific publications. Some of the databases also give the basis for measurements of citations, used as indicators of the quality of publications.



It will be noted that after an initial period of rapid expansion to a stable growth rate, the number of abstracts increases exponentially, doubling in approximately 15 years.

Fig. 1 Cumulative number of abstracts in various scientific fields, from the beginning of the abstract service to given data [1960]. From *Little Science, Big Science*, by Derek J. de Solla Price. Columbia Paperback Edition 1965. Copyright © 1963 Columbia University Press. Reprinted with permission of the publisher

In the present study we investigate the growth rate of science from 1907 to 2007. The study is based on information from databases for scientific publications and on growth data recorded in the literature. Using these data we have obtained time series from the beginning of the 20th century to 2007 with the best coverage from 1970 to 2005. The data give information about changes in the growth rate of science and permit a discussion about the internal and external causes of the observed changes.

The data have also been used to establish the coverage provided over time by the different databases. The dominant databases used in R&D statistics are Science Citation Index/Science Citation Index Expanded (SCI/SCIE) (SCIE is the online version of SCI), Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI). Together with other databases these databases are included in the Web of Science (WoS)

provided by Thomson Reuters, USA (Thomson Reuters 2008a). Of special interest is Conference Proceedings Citation Index (CPCI) (Thomson Reuters 2008b), partially overlapping with SCI/SCIE (Bar-Ilan 2009). It is necessary to specify the databases included in a search on WoS. In our work special attention has been paid to the coverage of SCI and SSCI.

One of the products from Thomson Reuters is National Science Indicators. This product is based solely on SCI/SCIE, SSCI and AHCI (Regina Fitzpatrick, Thomson Reuters, personal communication). Therefore, the coverage of this source is determined by the coverage of the citation databases.

The main focus of our work is on Natural and Technical sciences, not only because of the importance of these fields but also because publication patterns here are very different from those found in Social Science and Arts and Humanities.

We have not included Arts and Humanities, especially but certainly not only because of the importance of use of other languages than English (Archambault et al. 2005). An additional reason is the lack of suitable databases to compare with A&HCI.

Comparable problems are present for Social Sciences (Archambault et al. 2005). However, results obtained for Social Science using SSCI have validity and are therefore reported.

Based on the data from the databases included in our studies we address the following problems:

1. Is the growth rate of scientific publication declining?
2. Is the coverage by SCI and SSCI declining?
3. Is the role of conference proceedings increasing and is this reflected in the databases?

We are aware that many and important changes in publication methods are happening in the present years. These include open access archives, publications on the net, the increasing role of conference proceedings in many fields, the recent expansion of SCIE and SSCI (Testa 2008b) Conference Proceedings Citation Index from Thomson Reuters (Thomson Reuters 2009b) and the rapid expansion of Scopus and Google Scholar. Therefore, extrapolation from our results up to 2007 can not be made. However, vast amounts of bibliometric studies and scientometric studies is depending on publication numbers up to 2007 and will be so for a long time ahead.

We are also aware that counting of publications is treating all publications alike without regard to their widely different values. This is the major problem in scientometrics: Can all publications be treated alike and can they be added to provide meaningful numbers? Mathematically all units with common denominators can be added but this does not answer the problem. Statistically it can be hoped (or assumed) that the differences will be neutralized when large data sets are used for addition. However, this can not be proven (Garbage in, garbage out) and does not provide a solution.

Citation studies may say something about the value of individual publications but there are large differences between fields and the number of references per publication is steadily increasing in all fields. The “value” of a publication is also changing with time (Ziman 1968).

Anyway, publications are added all over the world for scientometric purposes. The lack of answer to the major problem posed above is not a deficiency of our publication. Publication numbers are of interest and are used generally in scientometrics and research statistics. It is impossible to combine a system based on giving values to individual publications with a study of the growth rate of science.

Methodology

Chemical Abstracts

Annual data for the total number of records in Chemical Abstracts (Chemical Abstracts Service, American Chemical Society) are available in CAS Statistical Summary 1907–2007. The data include separate values for *papers*, *patents* and *books*. Conference proceedings are also covered in Chemical Abstracts but are included under the heading papers and there are no separate figures for the number of proceedings. The share of papers slowly increased until about 1950. Since then the share has been relatively constant around 80%.

Compendex

Annual data for the *Total Number of Records* in Compendex (Engineering Village, Elsevier Engineering Information) from 1870 to 2007 were obtained on the net using the year in question as the search term and restricting the search to the same year. Compendex covers not only scientific publications in engineering but also other engineering publications. Therefore, comparisons with the other databases must be made with reservations. The values for 2004–2007 differed significantly from values received directly from Compendex. However, the growth rates for the two series were nearly identical. The values from 1988 to 2001 also differed from those reported for Compendex by National Science Foundation (Hill et al. 2007; Appendix, Table 1) but again the growth rates for the two series were similar.

CSA, Cambridge Scientific Abstracts

Annual data from CSA, Cambridge Scientific Abstracts, have been collected for *Natural Science* from 1977 to 2007 and for *Technology* from 1960 to 2007. The data includes values for *All Types*, *Journals*, *Peer-Reviewed Journals* and *Conference Proceedings*. However there are data breaks in most series, partly due to changes in the databases used as basis for the compilations.

Inspec

Values for Inspec and the sections of Inspec, *Computers/Control Engineering*, *Electrical/Electrical Engineering*, *Manufacturing and Production Engineering* and *Physics*, published by The Institution of Engineering and Technology, Stevenage, Herts., U.K., have been found on the net. The database was searched using the year in question as the search term and restricting the search to the same year. Values were found for the *Total Number of Records* as well as for *Journal Articles*, *Conference Articles* and *Conference Proceedings*. Inspec Physics is a direct continuation of Physics Abstracts but the change from the value from Physics Abstracts for 1969 to the value from Inspec Physics in 1970 indicates a break in the series.

Data were also obtained directly from Inspec but only giving the *Total Number of Records* from all sources. The data were not identical with those found on the net. However, the numbers of total records found on the net for the period 1969–2005 were only 1.6% higher than those given by Inspec. For the sections *Computers/Control Engineering*, *Electrical/Electrical Engineering*, *Manufacturing and Production Engineering*

and Physics, the corresponding values were 3.0, 3.1, 0.2 and 6.0%. For 2007 the differences were larger, probably because of the different dates for obtaining the values (July 8th, 2008, from Inspec, December 7th, 2008, from the net). The yearly values from 1969 to 2004 found on the net for Manufacturing and Production Engineering were identical with those obtained directly from Inspec.

We have chosen to use the Inspec data derived directly from the net because they included values both for the *Total Number of Records* and for *Journal Articles*, *Conference Articles* and *Conference Proceedings*.

LNCS

Data from LNCS (Lecture Notes in Computer Science, Springer Verlag) from 1940 to 2007 were found on the net, using the letter a as a search term and restricting the search to the year in question. Only values for *All Records* were obtained.

MathSciNet

Data from MathSciNet (American Mathematical Society) from 1907 to 2007 were obtained from the net, again using the year in question as the search term and restricting the search to the same year. Values were obtained for the categories *All Records*, *Journals*, *Proceedings* and *Books*. The values for Journals included not only articles but also book reviews and other items. Proceedings were only recorded from 1939.

Because publications in some cases are recorded both as books and as proceedings the sum of the values for Journals, Proceedings and Books are slightly higher than the values for All Records. Records from before 1940 do not provide complete coverage of the mathematical literature (personal communication from Drew Burton, American Mathematical Society).

Physics Abstracts

Annual data for the *number of records* in Physics Abstracts 1909-1969 were obtained from the published volumes. Distribution among books, journal articles, conference proceedings etc. could be obtained only by manual counting, an insuperable barrier.

PubMed Medline

Annual data for the *number of records* in PubMed Medline (National Library of Medicine, USA) 1959-2007 were obtained on the net. The data give no information about the distribution among books, journal articles, conference proceedings etc. The numbers for 1959-1965 are for a build-up period and the number for 2007 is unexpectedly high suggesting that this is not the final number.

SCI/SCIE

Annual data for the number of records in SCI (Thomson Reuters) from 1955 to 2007 have been obtained from the *Science Citation Index 2007 Guide*. Separate numbers are given for *Anonymous Source Items*, *Authored Source Items* and *Total Source Items*. From 1980 separate values are given for *Articles*, *Meeting Abstracts*, *Notes*, *News Items*, *Letters*, *Editorial Material*, *Reviews*, *Corrections*, *Discussions*, *Book Reviews*, *Biographical Items*,

Chronologies, Bibliographies, and Reprints. For the data from 1955 to 1964 a sum of 562 is given for source publications (journals). From 1965 to 1969 the number of source publications increased from 1,146 to 2,180, indicating a build-up period. In our comparisons and graphs we have only used the values from 1970 to 2005.

Science Citation Index Expanded (SCIE) is the online version of SCI. SCIE covers more than 6,650 journals across 150 disciplines. However, information about the number of publications recorded for each year is not available. Furthermore, when new journals are included also articles from previous years are added to SCIE. This means that the numbers of records for previous years in SCIE are changing with time and therefore cannot be used for time series. It also means that the coverage of SCIE is increasing continually as more and more databases are acquired by Thomson Reuters and included into SCIE.

SSCI

Annual data for the number of records in SSCI (Thomson Reuters) from 1966 to 2006 have been obtained from the *Social Sciences Citation Index 2005 Guide*. For 1966–1968 only the sum of the values for the 3 years are recorded. Separate numbers are given for *Anonymous Source Items*, *Authored Source Items* and *Total Source Items*. Numbers are given for both Selectively Covered Source Journals and Fully Covered Source Journals. From 1980 separate values are given for *Articles*, *Book Reviews*, *Letters*, *Editorials*, *Meeting Abstracts*, *Notes*, *Reviews*, *Corrections*, *Discussions*, *Biographical Items*, and *Chronologies*. We have only used the values from 1969 to 2005.

There is overlap between SSCI and SCI. This overlap is eliminated in the Web of Science (WoS).

Scopus

Annual data from 1997 to 2006 have been obtained by searching for the year in question and a* OR b* ... OR z* OR 0* ... OR 9*. Identical results were obtained using the advanced search function and searching for records after a fixed year (PUBYEAR AFT [digits for the year chosen]). The differences from year to year were the numbers for each year. Scopus is including new databases and journals backwards regularly. Therefore the results obtained depend on the time of acquisition. Scopus was searched on November 23rd 12, 2009. Separate numbers were obtained for *All Records*, *Articles and Reviews*, *Conference Papers*, *Conference Reviews*, *Letters and Notes*. The numbers for Conference Reviews were however insignificant compared with those obtained for the first other groups. Search was done for the four main groups, *Life Sciences*, *Health Sciences*, *Physical Sciences*, and *Social Sciences and Humanities* and for the combination of Life Sciences, Health Sciences and Physical Sciences. There are overlaps between the different groups and therefore the total numbers for Scopus are smaller than the sums of numbers for the groups. The advanced search function made it possible to search for smaller fields than the broad fields listed above. We have used this method to determine the percentage of articles and of conference proceedings in total records for the fields of computer science and engineering sciences in 2004 and 2004–2009. This search was performed on December 14th, 2009.

In our analysis we have used the numbers of *All Records*, including both authored and anonymous source items. We have compared “Papers” with “Articles and Reviews”, “Journals”, “Journal Papers”, “Journal Articles” and “Articles + Letters + Notes + Reviews” and use the common term *Journal Articles*. We have made this decision because it was the only way to obtain comparable results for the many databases studied. Each

database has its own system. It is not a question about what we think about articles, letters, notes and reviews, it is a question about what the database providers think and do.

We are aware that “letters” are used to name different types of publications in different journals. One remarkable case is the journal *Nature*. Here the major part of the publications is designated letters. This is just one example indicating that letters must be taken seriously. We are also aware that reviews only seldom contain reports of original research. However, if the value of a scientific publication shall influence its inclusion among publications reviews must be included. Some reviews have been very important in the development of science and reviews generally have more impact and receive more citations than articles.

We have not used the distinction between “Journals” and “Peer-reviewed Journals”, since the change of status for a journal does not provide information about publication activity. Furthermore peer review was not institutionalized before in the 1960s or 1970s. Proceedings of the National Academy of Sciences, USA, only recently introduced formal peer review. Thus, in all Figures and Tables we are using data given in the databases for all journal publications, also when data for peer-reviewed journals have been available. We have compared “Conference Proceedings” with “Conference Contributions”, “Conference Articles + Conference Proceedings”, “Conference Papers” and “Meeting Abstracts” and use the common term *Conference Contributions*.

We will discuss the role of Conference Contributions in “[Discussion, Conference contributions](#)”.

Our data for Social Sciences are restricted and permit only few conclusions.

Data for the number of journals covered by SCIE and SSCI have been obtained from the Web of Science.

Data for the number of journals covered by Scopus have been obtained from the home page of Scopus. Time series have been used to calculate annual growth rates and doubling times. Exponential growth has been studied using logarithmic display of time series. Linear regression has been used to calculate annual growth rates with standard errors and doubling times. Double sided tests have been used to calculate *P*-values for the difference between time series for different databases.

Results

Figure 2 gives a semi logarithmic presentation of the cumulative number of the total number of abstracts, the number of abstracts of papers and the number of abstracts of patents in Chemical Abstracts from 1907 to 2007, the total number of records in Compendex from 1907 to 2007, the total number of abstracts, the abstracts from journals and the abstracts of proceedings in MathSciNet from 1907 to 2007, the number of abstracts in Physics Abstracts (All Records) from 1909 to 1969 and the number of Abstracts (All Records) in Inspec Physics from 1969 to 2007. The graphs representing the total number of abstracts and covering the period from 1907 to 1960 are similar to Price’s classical figure in *Little Science, Big Science* (Fig. 1). Price interpreted the steep beginning of the curves as “an initial expansion to a stable growth rate” but of course the correct mathematical description is that for a curve giving cumulative values for exponential growth the slope is decreasing continually from a large initial value to a small annual growth rate. Price concluded from his data that the doubling period for science was about 15 years, corresponding to an annual growth rate of 4.73%. However, if the annual growth rate for exponential growth is 4.73% then the growth rate from year 53–54 (1959–1960) on the curve for cumulative values should be 5.18%. The slope observed for Chemical Abstracts

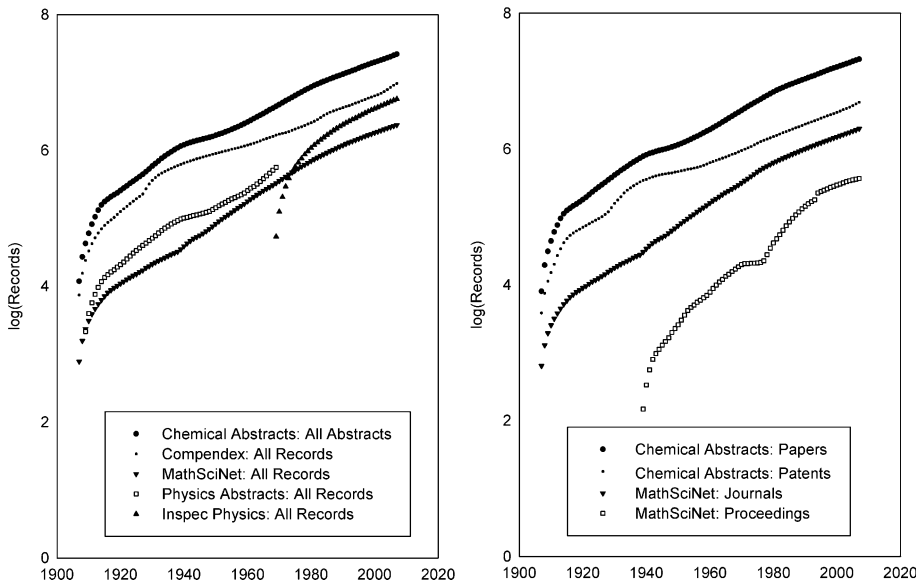


Fig. 2 Cumulative number of records for nine databases 1907–2007 (semi logarithmic scale)

in the linear period from 1952 to 1960 visible on Price's curve is about 1.048. However if the curve represents stable exponential growth the slope after 50 years is still significantly higher than a slope recording the exponential growth rate. Therefore the curve indicates an annual growth rate less than 4.2% and a doubling time higher than 15 years. Price used the number of All Records in Chemical Abstracts, not only the number of Journal Articles. If he had used only the number of journal articles he would have observed a growth rate approaching 9% per year and a doubling time about 9 years. The data for Physics Abstracts for the linear period from 1948 to 1960 visible on Price's curve indicate an annual growth rate less than 7.5% and a doubling time higher than 10 years. Data have not been available for Biological Abstracts to compare with the curve for Biological Abstracts in Price's figure. However, visual inspection of Price's figure also indicates a somewhat higher growth rate for the linear period from 1949 to 1960 (close to 6% per year, doubling time about 13 years). The effects of the two world wars are barely visible on the curves. Price mentioned a small decline in growth rate during World War II, but this can only be observed for Chemical Abstracts and Physics Abstracts (Price 1963, p. 10, 17).

In Fig. 3 the same numbers are presented. However, this figure records the number of abstracts for each year instead of the cumulative numbers. Again, the data are represented on a semi logarithmic scale. The straight lines represent a doubling time of 15 years (annual growth rate 4.73%).

Figure 3 gives more detailed information than Fig. 2. This is because information is lost in the integration giving the cumulative values. The curves for Chemical Abstracts reveal some interesting features. The negative effect of the two world wars and the extremely fast growth after the wars is clearly visible. Also the stagnation in the 1930s caused by the economic crisis from 1929 is clearly visible. This demonstrates that differences in growth rates can be explained and comprehended.

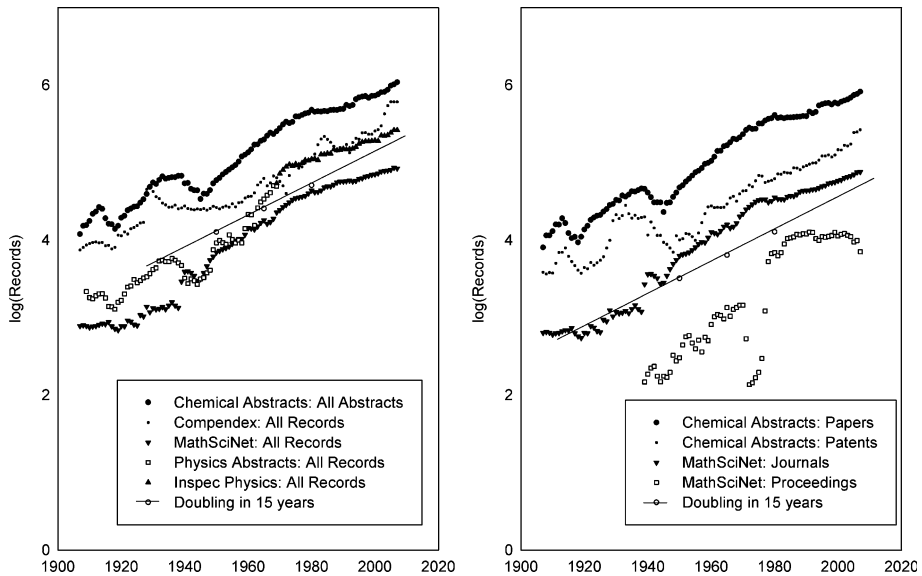


Fig. 3 Number of records for nine databases 1907–2007 (semi logarithmic scale)

In the period from 1974 to 1990 there is also a clear decline in the growth rate. This is followed by an increase in the period from 1990 to 2007 but the high values from before 1974 have not been reached again. Six different growth periods, 1907–1914, 1920–1930, 1930–1939, 1945–1974, 1974–1990, and 1990–2007, can be observed.

For Compendex the curve in Fig. 3 is very irregular, not permitting any conclusions. The curve for Compendex in Fig. 4 indicates general agreement with the other databases.

For MathSciNet there is a very high growth rate immediately after the end of World War II. The growth rate is still high up to the 1980s. At the end of the 1980s the growth rate has fallen to a very low level.

The data for Physics Abstracts show three periods, 1920–1930, 1930–1939, and 1945–1969, corresponding to the periods found for Chemical Abstracts. The data from Inspec Physics show a stable growth from 1971 to 2007 but the rate is much slower than that recorded in Physics Abstracts in the preceding period. This slow down corresponds with that found for chemistry.

Table 1 presents the growth rates and doubling times for the periods described above for Chemical Abstracts, Compendex, MathSciNet, Physics Abstracts and Inspec Physics. Slopes on the logarithmic scale and standard errors are included in the Table.

Figures 2 and 3 and Table 1 corroborate Price's work based on Biological Abstracts, Chemical Abstracts, Mathematical Reviews and Physics Abstracts although our analysis indicates a slightly lower growth rate for the period up to 1960 than that given by Price. The data for Chemical Abstracts also indicate that the growth in publication numbers has continued until 2007. However, the growth rate has not been stable. The growth in numbers of Journal Articles has declined significantly since 1974. The data for Physics Abstracts reflect the dramatic increase in growth from the end of World War II.

Figure 4 displays the graphs for All Records from 1970 to 2007 for Chemical Abstracts, Compendex, CSA Natural Science, CSA Technology, Inspec All Sources, Inspec

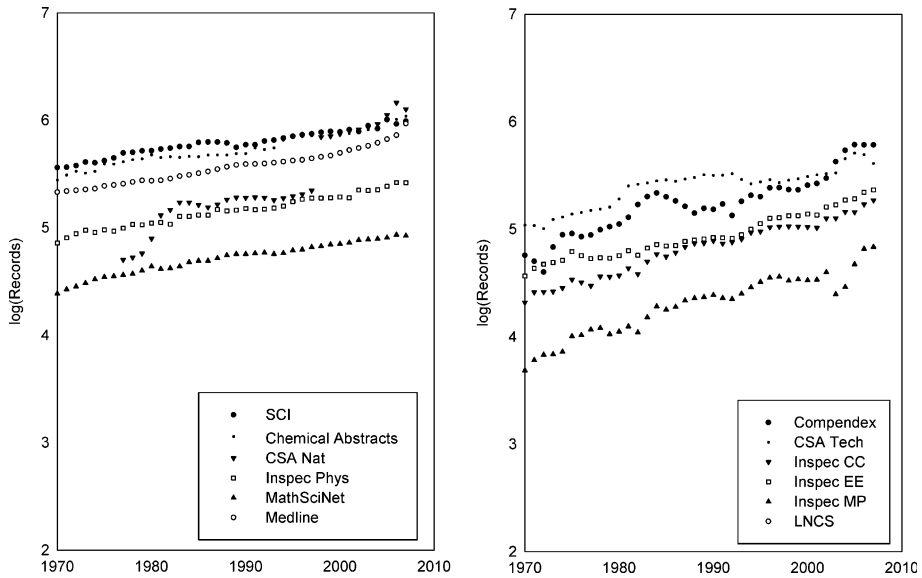


Fig. 4 Number of records for twelve databases 1970–2007 (semi logarithmic scale)

Electrical/Electrical Engineering, Inspec Computers/Control Engineering, Inspec Manufacturing and Production Engineering, Inspec Physics, LNCS, MathSciNet, Medline and SCI.

Figure 5 displays the graphs for Journal Articles from 1980 to 2007 for Chemical Abstracts, CSA Natural Science, CSA Technology, Inspec All Sources, Inspec Electrical/Electrical Engineering, Inspec Computers/Control Engineering, Inspec Manufacturing and Production Engineering, Inspec Physics, MathSciNet, and SCI.

Figure 6 displays the graphs for Conference Contributions from 1980 to 2007 for CSA Natural Science, CSA Technology, Inspec All Sources, Inspec Electrical/Electrical Engineering, Inspec Computers/Control Engineering, Inspec Manufacturing and Production Engineering, Inspec Physics, MathSciNet, and SCI.

The data obtained from Scopus are not recorded in Fig. 6 because they did not permit reliable calculations of growth rates. In all cases they showed nearly stagnation from 1997 to about 2002 and fast growth from about 2002 to 2006.

In Tables 2, 3 and 4 we present data from 1997 to 2006 derived from all the databases used except SSCI, Scopus and Inspec Manufacturing and Production Engineering. For Scopus and Inspec Manufacturing and Production Engineering the records fluctuated too much to permit reliable analysis. Probably, for many of the databases the values for 2007 are not final. 1997 has been chosen as the starting year because this is the first year with reliable data from all the databases included in the tables. Furthermore, the development in the most recent period is the most interesting, especially in connection with R&D statistics.

Tables 2, 3 and 4 indicate annual growth rates between 2.7 and 13.5% per year for the period 1997–2006 for All Records, between 2.2 and 9.0% per year for Journal Articles and between 1.6 and 14.0% per year for Conference Contributions. There are two possible explanations for this wide range. The first is that some of the databases increase or decrease coverage in their field. The second is that publication activity is growing with different rates in different fields.

Table 1 Growth rates for records in five databases 1907–2007

Period	All records		Journal articles		Patents	
	Slope on the logarithmic scale \pm standard error	Annual growth rate, %/ doubling time, years	Slope on the logarithmic scale \pm standard error	Annual growth rate, %/ doubling time, years	Slope on the logarithmic scale \pm standard error	Annual growth rate, %/ doubling time, years
Chemical abstracts						
1907–2007	0.019 \pm 0.0004	4.6/16	0.020 \pm 0.0004	4.8/15	0.017 \pm 0.0006	4.0/18
1907–1914	0.049 \pm 0.0051	12.0/6.1	0.045 \pm 0.0078	11.0/6.6	0.053 \pm 0.0091	13.1/5.6
1920–1930	0.043 \pm 0.0033	10.3/7.0	0.035 \pm 0.0017	8.4/8.6	0.066 \pm 0.011	16.4/4.6
1930–1939	0.010 \pm 0.0023	2.3/30	0.018 \pm 0.0018	4.2/17	–0.0048 \pm 0.0057	–1.1/–
1945–1974	0.035 \pm 0.0007	8.4/8.6	0.037 \pm 0.0009	8.8/8.2	0.031 \pm 0.0015	7.4/9.8
1974–1990	0.075 \pm 0.0011	1.7/40	0.0069 \pm 0.0014	1.6/44	0.012 \pm 0.0017	2.8/25
1990–2007	0.018 \pm 0.0011	4.1/17	0.015 \pm 0.0012	3.6/20	0.026 \pm 0.0020	6.2/12
Compendex						
1907–2007	0.017 \pm 0.0005	3.9/18				
MathSciNet						
1907–2007	0.025 \pm 0.0005	5.9/12	0.025 \pm 0.0005	5.9/12		
1907–1960	0.026 \pm 0.0012	6.1/12	0.027 \pm 0.0012	6.3/11		
1950–1984	0.026 \pm 0.0006	6.2/11	0.025 \pm 0.0009	6.0/12	0.029 \pm 0.0074	6.9/10
1984–2007	0.011 \pm 0.0003	2.5/28	0.014 \pm 0.0003	3.3/22	–0.0029 \pm 0.0015	–0.1/–
Physics abstracts						
1909–1960	0.017 \pm 0.0014	3.9/18				
Inspec physics						
1974–2004	0.013 \pm 0.0004	3.0/23	0.0092 \pm 0.0004	2.1/33	0.027 \pm 0.0012	6.4/11

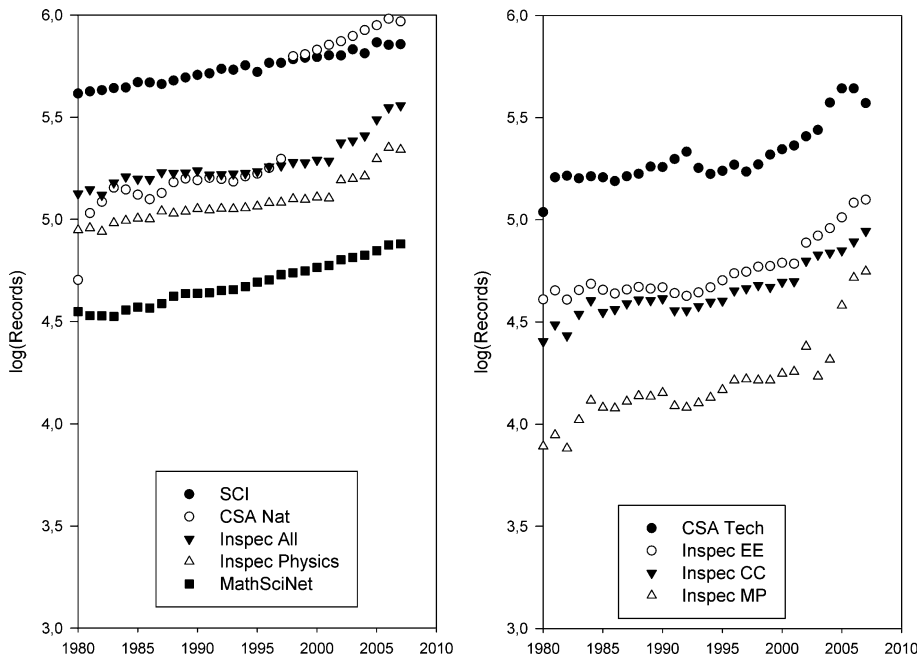


Fig. 5 Number of records for journal articles for nine databases 1980–2007 (semi logarithmic scale)

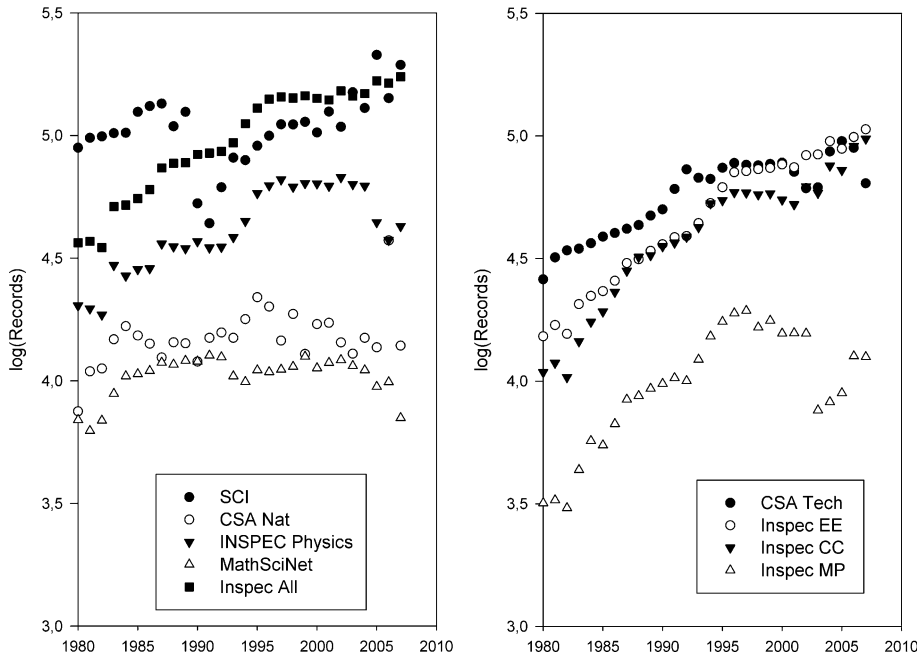


Fig. 6 Number of records for conference proceedings for nine databases 1980–2007 (semi logarithmic scale)

Table 2 Growth rates for all records in twelve databases 1997–2006

	Slope on the logarithmic scale \pm standard error	<i>P</i> -values declining the hypothesis “No difference between the database and SCI (double sided test)”	Annual growth rate, %/doubling time, years	Rank
Chemical abstracts	0.018 ± 0.0023	0.044	4.3/16	8
Compendex	0.055 ± 0.0073	<0.01	13.5/5.5	1
CSA, natural science. Only 1998–2004	0.020 ± 0.0011	<0.01	4.7/15	7
CSA, technology	0.031 ± 0.0043	<0.01	7.5/9.6	2
Inspec, all sources	0.021 ± 0.0031	0.017	5.0/14	6
Inspec computers/ control engineering	0.023 ± 0.0038	0.011	5.5/13	5
Inspec electrical/electrical engineering	0.026 ± 0.0028	<0.01	6.2/12	3
Inspec physics	0.016 ± 0.0022	0.17	3.8/18	10
LNCS. Only 1997–2006	0.018 ± 0.0017	0.041	4.2/17	9
MathSciNet	0.012 ± 0.00062	0.90	2.8/25	11
PubMed medline	0.024 ± 0.0012	<0.01	5.6/13	4
SCI	0.012 ± 0.0025	–	2.7/26	12

Table 3 Growth rates for journal articles in nine databases 1997–2006

	Slope on the logarithmic scale \pm standard error	<i>P</i> -values for the difference between the database and SCI (double sided test)	Annual growth rate, %/doubling time, years	Rank
Chemical abstracts	0.014 ± 0.019	0.040	3.4/21	8
CSA, natural science. Only 1998–2005	0.022 ± 0.00089	$\ll 0.01$	5.3/13	5
CSA, technology	0.048 ± 0.0040	$\ll 0.01$	7.5/9.6	2
Inspec, all sources	0.031 ± 0.0040	$\ll 0.01$	7.3/9.8	3
Inspec computers/ control engineering	0.028 ± 0.0030	$\ll 0.01$	6.6/11	4
Inspec electrical/ electrical engineering	0.037 ± 0.0039	$\ll 0.01$	9.0/8.0	1
Inspec physics	0.0028 ± 0.0030	<0.01	3.8/18	6
MathSciNet	0.016 ± 0.00078	$\ll 0.01$	3.7/19	7
SCI	0.0096 ± 0.0014	–	2.2/31	9

SCI has the lowest growth rate for All Records and for Journal Articles but the highest growth rate for Conference Contributions. When considering the high growth rate for Conference Contributions it must be taken into account that the share of Conference Contributions in All Records is low (see Table 5).

Table 4 shows that the importance of Conference Contributions differs between fields. The share in All Records of Conference Contributions has grown through the whole period covered for CSA, Cambridge Scientific Abstracts, Technology, Inspec Physics, Inspec

Table 4 Growth rates for conference contributions for five databases 1997–2006

	Slope on the logarithmic scale \pm standard error	<i>P</i> -values for the difference between the database and SCI (double sided test)	Annual growth rate, %/doubling time, years	Rank
CSA, technology	0.0073 \pm 0.0069	0.32	1.7/41	4
Inspec, all sources	0.0067 \pm 0.0020	0.010	1.6/45	5
Inspec computers/ control engineering	0.019 \pm 0.0055	<0.01	4.5/16	2
Inspec electrical/ electronical engineering	0.015 \pm 0.0020	\ll 0.01	3.6/20	3
SCI	0.057 \pm 0.015	–	14.0/5.3	1

Computers/ Control Engineering and Inspec Electrical/Electronical Engineering. The data for Inspec Manufacturing and Production Engineering shows a steady growth up to a 54% share in 1997, but thereafter a decline to about 18% in 2005. We have no explanation for this recent decline.

Table 5 gives the number of records for All Records, Journal Articles and Conference Contributions in 2004 for all the databases studied except SSCI. There is overlapping between the databases and thus the numbers cannot be added. However, it is remarkable that the number of records in SCI is lower than the numbers in Chemical Abstracts, CSA, Cambridge Scientific Abstracts, Natural Science, and Scopus, Life, Health and Physical Sciences combined and only slightly higher than the numbers in Medline and Scopus Physical Sciences. The table also reports the shares of All Records for articles and for conference proceedings.

Table 6 records for 1997–2006 the shares of Journal Articles and Conference Contributions in All Records for all databases investigated.

In SSCI there are substantial numbers of records for both reviews and editorial matter. Therefore, the sum of the shares for Journal Articles and Conference Contributions is only 68%.

The growth rate for SSCI for the period 1987–2006 has been found to be 1.6% per year (slope on a logarithmic scale 0.0069, Standard Error 0.0009) for All Sources and 2.0% per year (slope on a logarithmic scale 0.0081, Standard Error 0.0008) for Journal Articles. The corresponding doubling times are 44 and 37 years. The total number of All Records from 2000 to 2006 is 1,053,571. The values for Conference Contributions are too scattered to permit statistical analysis.

For comparison the growth rates for Scopus, Social Sciences and Humanities have for the period 1997–2006 been estimated to be 9% per year for All Records and 7% for Journal Articles.

Table 7 records the number of journals covered by SCIE and SSCI for the period from 1998 to 2009 (Thomson Reuters 2009a). The values for SCI for 1964 and 1972 have been found in (Garfield 1972) and the value for 1997 in (Zitt et al. 2003). Values for 2009 have been presented more recently (Thomson Reuters 2009b).

The search in Scopus for the fields Computer Science and Engineering to provide information about the relative roles of articles and conference proceedings have given the results displayed in Table 8.

Table 5 Number of records in databases 2004

Database	All records, numbers	Journal Articles		Conference contributions	
		Numbers	% of all records	Numbers	% of all records
Chemical abstracts	865,066	685,796	68.2	–	–
CSA natural science	917,780	844,273	92.0	14,960	1.6
CSA technology	452,744	374,333	82.7	86,401	19.1
Compendex	541,192				
Inspec, all sources	421,865	256,339	60.8	162,540	38.5
Inspec computers/control engineering	144,786	68,895	47.6	74,447	51.4
Inspec electrical/electrical engineering	186,421	90,969	48.8	93,944	50.4
Inspec physics	225,293	162,426	72.1	61,359	27.2
MathSciNet	78,829	66,761	85.0	11,046	14.0
PubMed medline	614,126	–	–	–	–
SCI	835,126	593,797	71.1	129,516	15.5
Scopus, life, health and physical sciences combined	1,415,911	1,089,135	76.9	260,576	18.4
Scopus, life sciences	398,497	361,105	90.6	19,676	4.9
Scopus, health sciences	434,215	381,220	87.8	21,424	4.9
Scopus, physical sciences	806,101	548,232	68.0	230,104	28.5
Scopus, social sciences and humanities	171,535	152,337	88.8	6,327	3.7
SSCI	133,643	81,324	60.9	10,306	7.7

Table 6 The share of journal articles and conference contributions in all records 1997–2006

Database	Share of journal articles (%)	Share of conference contributions (%)
CSA technology	78	21
CSA natural science	85	2
Inspec all records	61	39
Inspec physics	72	27
Inspec computers and control engineering	48	52
Inspec electrical/electronic engineering	49	51
Inspec manufacturing and production engineering	63	36
MathSciNet	84	15
SCI	78	16
Scopus, life, health and physical sciences combined	82	14
Scopus, life sciences	90	6
Scopus, health sciences	88	5
Scopus, physical sciences	76	21
Scopus, social sciences and humanities	91	4
SSCI	60	8

Table 7 Number of journals covered in SCIE and SSCI

Year	Number of journals covered					
	SCI	SSCI	Scopus			
			All fields combined	Life sciences	Health sciences	Physical sciences and humanities
1964	600					
1972	2,400					
1997	5,471					
1998	5,467	1,697				
1999	5,550	1,699				
2000	5,686	1,697				
2001	5,752	1,682				
2002	5,876	1,709				
2003	5,907	1,714				
2004	5,969	1,712				
2005	6,088	1,747				
2006	6,166	1,768				
2007	6,426	1,866				
2008	6,650	1,950				
2009	8,060	2,697	16,500	>4,300	>6,800	>7,200 >5,300

Table 8 The shares of articles and conference proceedings in 2004 and 2004–2009 according to Scopus

	Computer science per cent of all records		Engineering sciences per cent of all records	
	Articles	Conference proceedings	Articles	Conference proceedings
2004	34.9	49.2	25.9	62.9
2004–2009	29.7	61.9	35.6	54.0

Discussion

Analysis and interpretation of our results

It has been estimated that in 2006 about 1,350,000 articles were published in peer-reviewed journals (Björk et al 2008). The data suggest that the coverage in SCI is lower than in other databases and decreasing over time. It is also indicated that the coverage in SCI/SCIE is lower in high growth disciplines and in Conference Contributions than in well established fields like chemistry and physics. These indications are supported by complementary evidence from the literature. However, it must be remarked that SCI never has aimed at complete coverage, see below. The coverage of SCIE and SSCI has increased substantially in 2009, both by inclusion of more regional journals (Testa 2008a) and by general expansion/Thomson Reuters 2009b). The problems about the coverage of SCI/SCIE will be discussed again in “Fast- and slow-growing disciplines” and in the “Conclusion”.

The growth rate for SSCI is remarkably small. It is desirable to obtain supplementary information about the volume and growth rate of the publication activity in social sciences.

The growth rate of scientific publication and the growth rate of science

It is a common assumption that publications are *the* output of research. This is a simplistic understanding of the role of publication in science. Publication can just as well be seen as a (vital) part of the research process itself. Publications and citations constitute the scientific discourse (Ziman 1968; Mabe and Amin 2002; Crespi and Geuna 2008; Larsen et al. 2008). Nevertheless, the numbers of scientific publications and the growth rate for scientific publication generally are considered important science productivity or output indicators. The major producers of science indicators, the European Commission (EC), National Science Board/National Science Foundation (NSB/NSF, USA) and OECD all report publication numbers as output indicators (European Commission 2007; National Science Board 2008; OECD 2008). All base their data on SCI/SCIE, as do in fact virtually all others using publication number statistics. The data reported by NSB are nearly (but not completely) identical with those obtained directly from SCI/SCIE.

In 2008 NSB reported that the world S&E article output between 1995 and 2005 grew with an average annual rate of 2.3%, reaching 710,000 articles in 2005. This is based on the values for Articles + Letters + Notes + Reviews reported in SCI and in agreement with our results.

However, there are technical problems in counting publications (Gauffriau et al. 2007). In whole counting one credit is conferred to each country contributing to a publication. Whole counting involves a number of problems. Among these are that the numbers are non-additive, and therefore the publication number for a union of countries or for the world can be smaller than the sum of the publication numbers for the countries in the union or for the world. Indiscriminate use of whole counting leads to double counting. On the other hand, whole counting provides valuable information about the extent of scientific cooperation. In whole-normalized counting (fractional counting) 1 credit is divided equally between the countries contributing to a publication. Values obtained by whole-normalized counting are also non-additive. However, the values obtained by whole-normalized counting for large data sets are close to those obtained by complete-normalized counting (Gauffriau et al. 2008). In complete-normalized counting 1 credit is divided between the countries contributing to a publication in proportion to the number of institutions from each country contributing to the publication. Numbers obtained by complete-normalized counting are additive and can be used for calculating world shares. It is problematic that EC is using whole counting whereas NSB/NSF is using complete-normalized counting (National Science Board 2008). A publication by May (1997) with a high impact can serve as an example of the problems. A worldwide growth rate for scientific publication of 3.7% per year in the period 1981–1994 is reported. The real value, derived from SCI, is 2.3% per year. The incorrect figure must be due to the use of whole counting values and addition of non-additive numbers. The problems due to the use of different counting methods are also disclosed in comparisons of publication output between EU and USA. The use of whole counting shows a fast growth rate for EU-27 from 1981 to 2004 and a significant growth rate for USA from 1981 to 1995 but subsequently nearly no growth. Complete-normalized counting shows that the growth, both for EU and for USA, stopped completely in the period from 2000 to 2004 (Larsen et al. 2008). The counting problems are caused by scientific cooperation. If there was no scientific cooperation there would be no counting problems.

Mabe and Amin (2002) have given a precise description of the increasing extent of scientific cooperation. The authors write about the information explosion and to the rhetorical question “Is more being published?” give the answer “Based on papers published per annum recorded by ISI, the answer has to be an emphatic ‘yes!’”. However, using data from ISI they have for the period 1954–1998 calculated the number of papers per authorship, the average annual co-authorship, and the number of papers per unique author. The number of papers per authorship corresponds to whole counting. The number of papers per unique author is based on the total number of papers and the total number of active authors identified in the databases (the method used to solve the problem about homonyms is not stated). The number of authors per paper has increased from about 1.8 to about 3.7 in the period studied. Correspondingly, the number of papers per authorship has increased from about 1.8 to about 3.9. On the other hand, the number of papers per unique author has decreased from 1 to 0.8. Therefore the “productivity” of scientists has been decreasing slowly.

A possible explanation for this decrease in productivity is that in some disciplines a publication demands more and more work. Another possibility is that an increasing share of scientific publication consists of Conference Contributions not covered by the databases and publications presented for example on home pages or in open archives and again not covered by the databases. Anyway, Mabe and Amin conclude that “further analysis shows that the idea that scientists are slicing up their research into “least-publishable units” (or that “salami-style” publishing practices are occurring) appears to be unfounded.”

Mabe and Amin (2001) refer to National Science Foundation’s *Science and Engineering Indicators 2000* reporting a 3.2% annual growth in research and development manpower for a selection of six countries over the period 1981 to 1995. They write that data for the rest of the world are hard to obtain, but that a figure of around 3–3.5% is not unlikely for the world as a whole. In continuation they note that article growth in ISI databases has also been estimated at 3.5% in the period from 1981 to 1995. This is however not in agreement with our analysis of SCI data where we find a growth rate of 2.0% for all source items and 2.2% for Articles + Letters + Notes + Reviews. This again indicates that the “productivity” of science is decreasing when measured as the ratio between the number of traditional scientific publications and the scientific manpower.

Crespi and Geuna (2008) have discussed the output of scientific research and developed a model for relating the input into science to the output of science. They are aware that science produces several research outputs, classified into three broadly defined categories: (1) new knowledge; (2) highly qualified human resources; and (3) new technologies and other forms of knowledge that can have a socioeconomic impact. Their study is focused on the determinants of the first type of research output. There are no direct measures of new knowledge, but previous studies have used a variety of proxies. As proxies for the output of science they use published papers and citations obtained from the Thomson Reuters National Science Indicators (2002) database. However, they are aware of the shortcomings in these two indicators (see below).

The number of scientific journals

As mentioned in the “[Introduction](#)”, Price wrote that by 1950 the number of journals in existence sometime between 1650 and 1950 was about 60,000 and with the known growth rate the number would be about 1 million in year 2000 (Price 1961). This seems unrealistic but in 2002 it was reported that 905,090 ISSN numbers had been assigned to periodicals (Centre International de l’ISSN 2008). How many of these are scientific periodicals, how

many are in existence today, and are there periodicals not recorded in the international databases?

In 1981 it was reported that there were about 43,000 scientific periodicals in the British Library Lending Division (BLLD) and that BLLD attempted exhaustive coverage of the world's scientific literature with a stated policy of subscribing to any scientific periodical requested if it had scientific merit (Carpenter and Narin 1982).

The question has been taken up by Mabe and Amin (2001). Based on *Ulrich's International Periodicals Directory on CD-ROM*, they give a graphical representation of the numbers of unrefereed academic journals, refereed academic journals and active, refereed academic journals from 1900 to 1996. The number of unrefereed academic journals is about 165,000 in 1996. The numbers for refereed academic journals and active, refereed academic journals are about 11,000 and 10,500 in 1995. The growth rate for active, refereed journals is given as 3.31% per year for the period 1978–1996. In a subsequent publication (Mabe and Amin 2002) it is stated with reference to the first publication, that there are about 14,000 peer-reviewed learned journals listed in Ulrich's Periodicals Database. No information is given about the year for which the value of 14,000 is valid. Even if it is the year of the publication 2002, 3.31% annual growth from 1995 to 2002 gives only 13,188 journals but no explanation is given for this discrepancy.

However, in a third publication (Mabe 2003) it is reported that the number of active, refereed academic/scholarly serials comes to 14,694 for 2001. This number is based on a search using *Ulrich's International Periodicals Directory on CD-ROM*, Summer 2001 Edition. It is stated that this number is noticeably lower than estimates given by other workers but almost certainly represents a more realistic number. In this publication an annual growth rate of 3.25% is given for the period from 1970 to the present time.

Harnad et al. (2004) stated with reference to Ulrich that about 24,000 peer-reviewed research journals existed worldwide.

On the other hand van Dalen and Klamer (2005) reported that according to Ulrich's International Serials Database in 2004 about 250,000 journals were being published, of which 21,000 were refereed. Again, Meho and Yang (2007) stated that approximately 22,500 active academic/scholarly, refereed journals were recorded in Ulrich's Periodicals Directory.

According to Björk et al. (2008) the number of peer-reviewed journals was 23,750 in the winter of 2007. This figure was based on a search of Ulrich's database.

Scopus (see "[Citations and differences in citations recorded by different search systems](#)") in 2008 covers 15,800 peer-reviewed journals from more than 4,000 international publishers.

To conclude, the number of serious scientific journals today most likely is about 24,000. This number includes all fields, that is all aspects of Natural Science, Social Science and Arts and Humanities. There is no reason to believe that the number includes conference proceedings, yearbooks and similar publications. The number is of course important in considerations about the coverage of the various databases (see below in "[Citations and differences in citations recorded by different search systems](#)"). For comparison SCIE covered 6,650 journals and SSCI 1,950 journals in 2008 (Björk et al. 2008).

It must however be added that the criterion for regarding a journal as a serious scientific journal is peer review. Peer review in its modern present form is only about 40 years old and is not standardized. Therefore, the distinction between peer-reviewed journals and journals without peer review is not precise. It is worthwhile mentioning that a systematic peer review for Nature was only introduced in 1966 when John Maddox was appointed

editor of this journal. Proceedings of the National Academy of Sciences introduced peer review only a few years ago.

Citations and differences in citations recorded by different search systems

Until a few years ago, when citation information was needed, the single most comprehensive source was the Web of Science including SCI and SSCI but recently two alternatives have become available.

Scopus was developed by Elsevier and launched in 2004 (Reed Elsevier 2008). In 2008 Scopus covers references in 15,800 peer-reviewed journals.

Google Scholar records all scientific publications made available on the net by publishers (Google 2008). A publication is recorded when the whole text is freely available but also if only a complete abstract is available. The data comes from other sources as well, for example freely available full text from preprint servers or personal websites.

A number of recent studies have compared the number of citations found and the overlap between the citations found using the three possibilities.

The use of Google Scholar as a citation source involves many problems (Meho 2006; Bar-Ilan 2008). But it has repeatedly been reported that more citations are found using Google Scholar than by using the two other sources and also that there is only a limited overlap between the citations found through Google Scholar and those found using the Web of Science (Meho 2006; Meho and Yang 2007; Bar-Ilan 2008; Kousha and Thelwall 2008; Vaughan and Shaw 2008, and references therein).

Meho and Yang (2007) have studied the citations found in 2006 for 1,457 scholarly works in the field of library science from the School of Library and Information Science at Indiana University-Bloomington and published in the period from 1970 to 2005. 2,023 citations of these publications in the period from 1996 to 2005 were found in the WoS, in Scopus 2,301 and in Google Scholar 4,181. There was a great deal of overlap between WoS and Scopus but Scopus missed about 20% of the citations caught in WoS whereas WoS missed about 30% of the citations caught in Scopus. There was restricted overlap between on the one side WoS and Scopus and on the other side Google Scholar. 60% of the citations caught in Google Scholar were missed by both WoS and Scopus whereas 40% of the citations caught in WoS and/or Scopus were missed by Google Scholar.

Kousha and Thelwall (2008) have reported a study involving the comparisons of citations in four different disciplines, biology, chemistry, physics and computers (In all fields only journals were included when giving open access and therefore accessible to Google Scholar as well as to WoS). The citations were collected in January 2006. From the data given in Table 1, page 280, it can be calculated that the ratios found for the four fields between citations found in Google Scholar and in WoS are 0.86, 0.42, 1.18 and 2.58. The citations common to WoS and Google Scholar represented 55, 30, 40 and 19% respectively of the total number of references. The dominant types of Google Scholar unique citing sources were journal papers (34.5%), conference/workshop papers (25.2%) and e-prints/preprints (22.8%). There were substantial disciplinary differences between types of citing documents in the four disciplines. In biology and chemistry 68, respectively 88.5% of the unique citations from Google Scholar were from journal papers. In contrast, in physics e-prints/preprints (47.7%) and in computer science conference/workshop papers (43.2%) were the major sources of unique citations in Google Scholar.

Vaughan and Shaw (2008) have studied the citations of 1,483 publications from American Library and Information Science Faculties. The citations were found in

December 2005 in the Web of Science and in the spring of 2006 in Google and Google Scholar. Correlations between Google and Google Scholar were high whereas WoS and web citation counts varied. Using Table 1 (page 323) in the publication it can be calculated that a total of about 3,700 citations were found on WoS whereas about 8,500 citations were found in Google Scholar. More citations were found on Google Scholar for all types of publications but whereas the ratio between Google Scholar citations and WoS citations were 8 and 6.4 for conference papers and open access articles the ratio was only 1.6 for publications in subscription journals.

Smith (2008) has investigated the citations found in Google Scholar for universities in New Zealand. There are no direct comparisons with WoS or SCOPUS but the conclusion is that Google Scholar provides good coverage of research based material on the Web.

Using WoS and SciFinder from Chemical Abstracts Service for a random sample of 15 chemists Whitley (2002) reported 3,234 citations in SciFinder, 2,913 in WoS. 58% of the citations were overlapping, 25% were unique for SciFinder and 17% were unique for WoS. For a second random sample of 16 chemists similar results were obtained.

According to Mabe (2003) the ISI journal set represents about 95% of all journal citations found in the ISI database. This conclusion is supported with a reference to Bradford's Law (Bradford 1950; Garfield 1972, 1979), a bibliometric version of the Pareto Law, often called the Matthew Principle: 'to him that hath shall be given' (Merton 1968, 1988). This indicates that citations found in SCI and SSCI are primarily based on the journals covered by these databases.

Bias in source selection and language barriers

When SCI and later SSCI were established it was the ambition to cover the most important part of the scientific literature but not to attempt complete coverage. This is based on the assumption that the significant scientific literature appears in a small core of journals in agreement with Bradford's Law (Garfield 1972, 1979). Journals were chosen by advisory boards of experts and by large scale citation analysis. The principle for selecting journals has been the same during the whole existence of the citation indexes. New journals are included in the databases if they are cited significantly by the journals already in the indexes and journals in the indexes are removed if their numbers of citations in the other journals in the indexes are declining below a certain threshold. A recent publication provides a detailed description of the procedure for selecting journals for the citation indexes (Testa 2008a).

From soon after the inception of SCI, it has been criticized for being biased toward papers in the English language and those from the United States (Shelton et al. 2009). As an example, MacRoberts and MacRoberts (1989) noted that SCI and SSCI covered about 10% of the scientific literature. The figure of 10% is not substantiated in the publication or in the references cited. However, it is clearly documented that English language journals and western science were over-represented; whereas small countries, non-western countries, and journals published in non-Roman scripts were under-represented. Thorough studies of the problems inherent in the choice of journals covered by SCI have been reported (van Leuwen et al. 2001; Zitt et al. 2003).

A study of public health research in Europe covered 210,433 publications found in SCI and SSCI (with exclusions of overlap). Of the publications 96.5% were published in English, 3.5% in a non-English language, with German as the most common. Therefore the dominance of journals with English language was clearly visible. It is difficult to make firm estimates about how many non-English valuable publications were missed but it is a reasonable conjecture that the number is substantial (Clarke et al. 2007).

Crespi and Geuna (2008) have recently reported a cross country analysis of scientific production. As mentioned above they state that the main source of the most commonly used two proxies for the output of science variables is the Thomson Reuters National Science Indicators (2002) database of published papers and citations. Among the shortcomings in this source is that the Thomson Reuters data are strongly affected by the disciplinary propensity to publish in international journals and that the ISI journal list is strongly biased towards journals published in English, which will lead to an underestimation of the research production of those countries where English is not the native language.

A special report from NSF in 2007 (Hill et al. 2007) contains a short discussion about the coverage of Thomson ISI Indexes. It is mentioned that “journals of regional or local importance may not be covered, which may be especially salient for research in engineering/technology, psychology, the social sciences, the health sciences, and the professional fields, as well as for nations with a small or applied science base. Thomson ISI covers non-English language journals, but only those that provide their article abstracts in English, which limits coverage of non-English language journals”. It is also stated that these indexes relative to other bibliometric databases cover a wider range of S&E fields and contain more complete data on the institutional affiliations of an article’s authors. For particular fields, however, other databases provide more complete coverage. Table 1 in the Appendix of the report presents publication numbers for USA and a number of other countries derived from Chemical Abstracts, Compendex, Inspec and PASCAL for the period 1987–2001. Publications have been assigned to publishing centre or country on the basis of the institutional address for the first author listed in the article. The values for the world can be obtained from the table by addition. According to Chemical Abstracts for 23.0% of the publications the first author was from USA. The values found were for Compendex 25.1%, for Inspec 22.7% and for PASCAL 29.0%. For comparison, the share for USA according to SCI is 30.5% (National Science Foundation 2006). This indicates that SCI is biased towards publications from USA to a higher degree than the other databases. Another possibility is that SCI is fair in its treatment of countries whereas the other databases are biased against USA; this is not a very likely proposition.

In a study of the publication activities of Australian universities Butler (2008) has calculated the coverage of WoS for all publications and for journal articles for publications from 1999 to 2001. The study was based on a comparison between publications recorded in WoS for all publications and for journal articles from 1999 to 2001 and a national compilation of publication activities in Australian universities in the same period. In WoS was found from 74 to 85% of the nationally recorded publications in biological, chemical and physical sciences. The coverage was better for journal articles, from 81 to 88%. For Medical and Health Sciences the coverage in WoS was slightly lower, 69.3 for all articles and for journal articles 73.7%. Again for Agriculture, Earth Sciences, Mathematical Sciences and Psychology the coverage in WoS for all publications was between 53 and 64%, for journal articles between 69 and 79%. For Economics, Engineering and Philosophy the coverage in WoS for all publications was between 24 and 38%, for journal articles between 37 and 71%. For Architecture, Computing, Education, History, Human Society, Journalism and Library, Language, Law, Management, Politics and Policy and The Arts the coverage in WoS was between 4 and 19% for all publications, between 6 and 49% for Journal Articles. These data clearly indicate that it is deeply problematic to depend on WoS for publication studies in Humanities and Social Sciences but also in Computing and Engineering.

In a recent study it is demonstrated that in Brazil the lack of skill in English is a significant barrier for publication in international journals and therefore for presence in WoS (Vasconcelos et al. 2009).

A convincing case has also been made that SSCI and AHCI are not well suited for rating the social sciences and humanities (Archambault et al. 2005).

As part of a response to such criticism Thomson Reuters has recently taken an initiative to increase the coverage of regional journals (Testa 2008b). However, the share of publications from the USA in journals newly added to SCI/SCIE is on average the same as the share in the “old” journals covered by SCI/SCIE. This indicates that if there is a bias in favour of USA, it has not changed in recent years (Shelton et al. 2009).

Conference contributions

Conference Proceedings have different roles in different scientific fields. As a generalisation it can be said that the role is smallest in the old and traditional disciplines and largest in the new and fast growing disciplines. In some fields conference proceedings are not considered as real publications, considered as abstracts and not subjected to peer review and are generally expected to be followed by real publications. The proceedings are not published with ISBN- or ISSN-numbers and not available on the net. In other fields conference proceedings provide the most important publication channel. Table 8 indicates that conference proceedings are much more important than journal articles in computer science and engineering sciences. In many fields conference contributions are subjected to meticulous peer review. A natural example in our context is the biannual conferences under the auspices of ISSI, The International Society for Scientometrics and Bibliometrics. In many engineering sciences the rejection rate for conference contribution is high. Conference proceedings are provided with ISBN- or ISSN-numbers and often available on the net or in printed form at the latest at the beginning of the conference.

Therefore, in a study of the growth rate of science and the coverage of databases it does not make sense to say no to conference proceedings. It makes sense to include them but be aware of their different roles in different fields when interpreting the results.

Table 4 shows that SCI has a relatively low share of Conference Contributions among the total records. There is however one exception, the complete coverage of “lecture notes in ...” series published by Springer, which publishes conference proceedings in computer science and mathematics in book form (Björk et al. 2008).

Thomson Reuters has covered conference proceedings from 1990 in ISI Proceedings with two sections, Science and Technology and Social Sciences and Humanities. However, these proceedings were not integrated in the WoS until 2008. Therefore, the proceedings recorded have not been used in scientometric studies based on SCI and SSCI.

In 2008 Thomson Reuters launched Conference Proceedings Citation Index, fully integrated into WoS and with coverage back to 1990 (Thomson Reuters 2008b). A combination of this new Index with SCIE and SSCI will give a better total coverage. However, if scientometric studies continue to be based solely on SCI and SSCI, the low coverage of conference proceedings there will still cause problems.

The weak coverage in WoS of Computer Science and Engineering Sciences mentioned in “[Bias in source selection and language barriers](#)” is probably caused by the low coverage of conference proceedings (Butler 2008).

The inclusion of conference proceedings in databases may cause double counting when nearly or completely identical results are first presented at a conference and later published in a journal article. Again, this is field dependent. In areas where conference proceedings

have great importance it is common that publication in proceedings is not followed up by publication in a journal. On the other hand, in areas where conference proceedings are of lesser importance they are often not covered by the databases.

Fast- and slow-growing disciplines

In “[Analysis and interpretation of our results](#)” the different growth rates for different scientific disciplines were discussed. There are indications that many of the traditional disciplines, including chemistry, mathematics and physics, are among the slowly growing disciplines, whereas there are high growth rates for new disciplines, including engineering sciences and computer science. Engineering sciences and computer science are disciplines where conference proceedings are important or even dominant. There has through the years been a discussion about and criticism of the coverage in SCI of computer science (Moed and Visser 2007). However, a special effort has been made recently to increase the coverage of computer science in SCI/SCIE, see “[Conference contributions](#)”.

Most recently (April 20, 2009), the database INSPEC with a stronger coverage of conference proceedings in the engineering sciences was integrated in the database Web of Science (UC Davis University Library Blogs 2009). The influence of this integration (double counting of conference proceedings and corresponding journal articles as well as better coverage of the literature) is yet to be studied.

Do the ISI journals represent a closed network?

SCI has been the dominant database for the counting of publications and citations. Because of the importance of the visibility obtained by publishing in journals covered by this database and because of the use of the counting values in many assessment exercises and evaluations, it has been important for individual scientists, research groups, institutions and countries to publish in the journals covered by this database. The Hirsch Index (Hirsch 2005) is one example of a science indicator derived from SCI. It is a reasonable conjecture that SCI has had great influence on the publishing behaviour among scientists and in science.

But the journals in SCI constitute a closed set. It is not easy for a new journal to gain entry. One way to do so is to publish papers bringing references to the journals already included. It is important to publish in English since English speaking authors and authors for whom English is the working language only rarely cite literature in other languages. It is also helpful to publish in journals in which most of the publications come from the major scientific countries. No scientist can read everything which may be of potential interest for his or her work. A choice is made and the choice is to select what is most easily available, what comes from well-known colleagues and what comes from well-known institutions and countries.

As mentioned above, Zitt et al. (2003) have made a detailed study of the problems inherent in the choice of journals covered by SCI.

All in all, it is best to get inside but it is not easy. A recent publication about the properties of “new” journals covered by SCI (new means included in 1995 or later) is of interest in this connection. On average the new journals had the same distribution of authors from different countries as the “old” journals (old means included before 1995). Therefore the new journals are not an open road for scientists from countries with a fast growth in publication activity (as for example the Asian Tigers, China, South Korea and Singapore). The new journals are just more of the same (Shelton et al. 2009).

The role of in-house publications, open access archives and other publications published on the net

These forms of publications are fast gaining in importance. The new publication forms may invalidate the use of publication numbers derived from the big databases in measurements of scientific productivity or output and of the growth rate of science. The effect cannot be determined by the data analysed by us. However, there is good reason to believe that a fundamental change in the publication landscape is underway.

Has the growth rate of science been declining?

Price in 1963 concluded that the annual growth rate of science measured by number of publications was about 4.7% (Price 1963). The annual growth rates of 3.7% for Chemical Abstracts for the period 1907–1960 and of 4.0% for Physics Abstracts for the period 1909–1960 given in Table 1 are lower (see Tables 1).

What has happened since then? Tables 2, 3 and 4 show a slower growth rate in the period 1997 to 2006 according to SCI, MathSciNet and Physics Abstracts. Most other databases indicate an annual growth rate above 4.7%. The same can be concluded for the period from 1960 to 1996 but long time series are only available for some of the databases used as basis for Tables 2, 3 and 4.

A tentative conclusion is that old, well established disciplines including mathematics and physics have had slower growth rates than new disciplines including computer science and engineering sciences but that the overall growth rate for science still has been at least 4.7% per year. However, the new publication channels, conference contributions, open archives and publications available on the net, for example in home pages, must be taken into account and may change this situation.

Conclusion

In the introduction three questions were asked.

The first question is whether the growth rate of scientific publication is declining? The answer is that traditional scientific publishing, that is publication in peer-reviewed journals, is still increasing although there are big differences between fields. There are no indications that the growth rate has decreased in the last 50 years. At the same time, publication using new channels, for example conference proceedings, open archives and home pages, is growing fast.

The second question is whether the coverage of SCI and SSCI declining?

It is clear from our results and the literature that the growth rate for SCI is smaller than for comparable databases, at least in the period studied. This means that SCI is covering a decreasing part of the traditional scientific literature. There are also clear indications that the coverage of SCI is especially low in some of the scientific areas with the highest growth rate, including computer science and engineering sciences.

The third question is whether the role of conference proceedings is increasing and whether this is reflected in the databases? The answer is that conference proceedings are especially important in scientific fields with high growth rates. However, the growth rates for conference proceedings generally are not higher than those found for Journal Articles. It is clear that the increasing importance of conference proceedings is only partially reflected in SCI.

It is problematic that SCI has been used and is still used as the dominant source for science indicators based on publication and citation numbers. SCI has nearly been in a monopoly situation. This monopoly is now being challenged by the new publication channels and by new sources for publication and citation counting. It is also a serious problem because a substantial amount of scientometric work and of R&D statistics has been done using a database which year for year has covered a smaller part of the scientific literature.

National Science Indicators is one of the products offered by Thomson Reuters. Since this product is based solely on SCI/SCIE, SSCI and AHCI the use of this product is problematic.

The recent expansion of SCIE and SSCI (Testa 2008b, Thomson Reuters 2009b) does not provide a solution to the problems. If new journals are included backward this means that previous scientometric studies based on SCIE and SSCI cannot be compared with studies using the current content of the databases (Hill et al. 2007). Of course, it is no solution to include new journals in current years without updating previous years.

Therefore, an expanding set of journals poses problems for trend analyses. On the other hand, working with a fixed set of journals is also posing problems. Because new research communities often spawn new journals to disseminate their research findings, a fixed journal set under-represents the types of research that were not already well established at the outset of the period. The longer the period being studied, the less adequate a fixed journal set becomes as a representation of the world's articles throughout the period (Hill et al. 2007, the section on methodological issues).

These conclusions may not be helpful. It is not clear what should be done in the future. A big and obvious question is also when and how the growth rate in science will decline. Simple logic tells us that this must happen long before the whole population of the world has turned into scientists. We don't know the answer. However, a conjecture is that the borderline between science and other endeavours in the modern, global society will become more and more blurred.

Acknowledgements We are indebted to Dr. Else Marie Bartels, Copenhagen University Library, Denmark, for help in database search, Daniela Bausano, Inspec, The Institution of Engineering and Technology, UK, for data about Inspec, Regina Fitzpatrick and Dr. Henry Small, Thomson Reuters, USA, for data about SCI and SSCI, Regina Fitzpatrick for information about National Science Indicators from Thomson Reuters, Wilfred T. Hastings, Aalborg, Denmark, for correction of the language, Ove Kähler, Elsevier, for information on search methods applicable to Scopus, Severin Olesen Larsen, Statens Serum Institut, Copenhagen, Denmark, for help in statistical analyses, professor Peter Laur, Aachen, Germany, for help in obtaining data for Chemical Abstracts, Eric Shively, Chemical Abstracts Service, USA, for information about the coverage of conference proceedings in Chemical Abstracts, and Finn Torben Sørensen, The Library, Faculty of Life Science, University of Copenhagen, Denmark, for help in collecting data for Medline. Support from the Carlsberg Foundation is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Archambault, É., Gagné, É.-T., Côté, G., Larivière, V., & Gingras, Y. (2005). Welcome to the linguistic warp zone: Benchmarking scientific output in the social sciences and humanities. In P. Ingwersen & B. Larsen, (Eds.), *Proceedings of the ISSI 2005 Conference*, Stockholm, July 24–28 (Vol. 1, pp. 149–158).

- Bar-Ilan, J. (2008). Which h-index?—a comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74, 257–271.
- Bar-Ilan, J. (2009). Web of Science with the conference proceedings citation indexes—the case of computer science. In B. Larsen & J. Leta (Eds.), *Proceedings of ISSI 2009—the 12th international conference of the international society for scientometrics and informetrics*, Rio de Janeiro, Brazil (Vol. 1, pp. 399–409). Published by BIREME/PAHO/WHO and Federal University of Rio de Janeiro, 2009. ISSN 2175-1935.
- Björk, B.-C., Roos, A., & Lauro, M. (2008). Global annual volume of peer reviewed scholarly articles and the share available via Open Access options. *Proceedings ELPUB2008 Conference on Electronic Publishing*. Toronto, Canada, June 2008, pp. 1–10. Retrieved December 9, 2008 from <http://elpub.scix.net>.
- Bradford, S. C. (1950). *Documentation*, Washington DC, Public Affairs Office. See: Bradford, S.C. (1953). *Documentation*, London: Crosby Lockwood & Sons.
- Butler, L. (2008). ICT assessment: Moving beyond journal outputs. *Scientometrics*, 74, 39–55.
- Carpenter, M. P., & Narin, F. (1982). The adequacy of the *Science Citation Index* (SCI) as an indicator of international scientific activity. *Journal of the American Society for Information Science*, 32, 430–439.
- Centre International de l'ISSN, 2008. Retrieved November 19, 2008 from <http://www.issn.org>.
- Clarke, A., Gatineau, M., Grimaud, O., Royer-Devaux, S., Wyn-Roberts, N., Le Bis, I., et al. (2007). A bibliometric overview of public health research in Europe. *European Journal of Public Health*, 17(Supplement 1), 43–49.
- Crespi, G. A., & Geuna, A. (2008). An empirical study of scientific production: A cross country analysis, 1981–2002. *Research Policy*, 37, 565–579.
- European Commission. (2007). *Towards a European research area. Science, Technology and Innovation. Key Figures 2007*. Retrieved May 8, 2008 from <http://ec.europa.eu>.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178, 472–479.
- Garfield, E. (1979). *Citation indexing: Its theory and application in science. Technology and humanities*. New York: Wiley.
- Gaufrriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, 73, 175–214.
- Gaufrriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparison of results of publication counting using different methods. *Scientometrics*, 77, 147–176.
- Google. (2008). *About Google Scholar, support for scholarly publishers*. Retrieved November 19, 2008 from <http://scholar.google.com>.
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30, 310–314.
- Hill, D., Rapoport, A. I., Lehming, R. F., & Bell, R. K. (2007). *Changing U.S. output of scientific articles: 1988–2003*. National Science Foundation, Division of Science Resources Statistics, NSF 07-320. Retrieved November 24, 2008 from <http://www.nsf.gov/statistics/nsf07320>.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific output. *Proceedings of the National Academy of Science of the United States of America*, 102, 16569–16572.
- Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four disciplines. *Scientometrics*, 74, 273–294.
- Larsen, P. O., Maye, I., & von Ins, M. (2008). Scientific output and impact: Relative positions of China, Europe, India, Japan and the USA. *The COLLNET Journal of Scientometrics and Information Management*, 2(2), 1–10.
- Larsen, P.O., & von Ins, M. (2009). The steady growth of scientific publication and the declining coverage provided by science citation index. In B. Larsen & J. Leta (Eds.), *Proceedings of ISSI 2009—the 12th international conference of the international society for scientometrics and informetrics*, Rio de Janeiro, Brazil (Vol. 2, pp. 597–606). Published by BIREME/PAHO/WHO and Federal University of Rio de Janeiro, 2009. ISSN 2175-1935.
- Mabe, M. (2003). The growth and number of journals. *Serials*, 16(2), 191–197.
- Mabe, M. A., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51, 147–162.
- Mabe, M. A., & Amin, M. (2002). Dr Jekyll and Dr Hyde: Author-reader asymmetries in scholarly publishing. *Aslib Proceedings*, 54, 149–175.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40, 342–349.
- May, R. (1997). The scientific wealth of nations. *Science*, 275, 793–796.
- Meho, L. I. (2006). The rise and rise of citation analysis. *Physics World (arXiv:physics/0701012v1)*, pp. 1–15.

- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIES faculty: Web of science versus scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58, 2105–2125.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159, 56–63.
- Merton, R. K. (1988). The Matthew Effect in science, II, Cumulative advantage and the symbolism of intellectual property. *Isis*, 79, 607–623.
- Moed, H. F., & Visser, M. S. (2007). *Developing bibliometric indicators of research performance in computer science: An exploratory study. Research Report to the Council for Physical Sciences of the Netherlands Organisation (NOW). CWTS Report 2007-1*. Retrieved January 9, 2008 from http://www.socialsciences.leidenuniv.nl/cwts/publications/nwo_indcompsciv.lsp.
- National Science Board. (2008). *Science and engineering indicators 2008*. Retrieved January 15, 2008 from <http://www.nsf.gov/statistics/seind08>.
- National Science Foundation. (2006). *Science and engineering Indicators 2006*.
- OECD. (2008). *Science, Technology and industry outlook*. Paris. ISBN 978-92-64-04991-8.
- Price, D. J. de S. (1961). *Science since Babylon*. New Haven, Connecticut: Yale University Press.
- Price, D. J. de S. (1963). *Little science. Big Science*. New York: Columbia University Press.
- Price, D. J. de S. (1965). The scientific foundations of science policy. *Nature*, 206, 233–238.
- Reed Elsevier. (2008). *Scopus*. Retrieved November 19, 2008 from <http://www.reed-elsevier.com>.
- Shelton, R. D., Foland, P., & Gorelsky, R. (2009). Do new SCI journals have a different national bias? *Scientometrics* (in press). doi:10.1007/s11192-009-0423-1.
- Smith, A. G. (2008). Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise. *Scientometrics*, 74, 309–316.
- Testa, J. (2008a). *The Thomson Scientific journal selection process*. Retrieved December 7, 2008 from http://www.thomsonreuters.com/business_units/scientific/free/essays/journalselection/ pp. 1–6.
- Testa, J. (2008b). *Regional content expansion in web of science: Opening borders to exploration* (pp. 1–3). Retrieved December 7, 2008 from http://www.thomsonreuters.com/business_units/scientific/free/essays/regionalcontent/.
- Thomson Reuters. (2008a). *ISI web of knowledge, science citation index expanded, social science citation index, web of science*. Retrieved January 15, 2009 from http://thomsonreuters.com/products_services/scientific/ISI_Web_of_Knowledge.
- Thomson Reuters. (2008b). *Conference proceedings citation index*. Retrieved January 15, 2009 from http://www.thomsonreuters.com/products_services/scientific/Conf_Proceedings_Citation_Index.
- Thomson Reuters. (2009a). *ISI web of knowledge, journal citation reports, science edition and social sciences edition*. Retrieved February 1, 2009.
- Thomson Reuters. (2009b). *ISI web of knowledge, journal citation reports, science edition and social sciences edition*. Retrieved July 30, 2009.
- UC Davis University Library Blogs, Department Blog, Physical Sciences & Engineering Library (2009), INSPEC moving to Web of Science site, <http://blogs.lib.ucdavis.edu/pse/2009/04/08/inspec-moving-t-web-of-science-site/>. Retrieved August 6, 2009.
- van Dalen, H. P., & Klamer, A. (2005). Is science a case of wasteful competition? *KYKLOS*, 58, 395–414.
- Van Leuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the *Science Citation Index* and its consequences for international comparisons of national research performance. *Scientometrics*, 51, 335–346.
- Vasconcelos, S. M. R., Sorenson, M. M., & Leta, S. (2009). A new input indicator for the assessment of science and technology research. *Scientometrics*, 80, 217–230.
- Vaughan, V., & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74, 317–330.
- Whitley, K. M. (2002). Analysis of SciFinder scholar and web of science citation searches. *Journal of the American Society for Information Science and Technology*, 53, 1210–1215.
- Ziman, J. (1968). *Public knowledge. An essay concerning the social dimension of science*. Cambridge: University Press.
- Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2003). Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation. *Scientometrics*, 56, 259–282.