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

Scientometrics,
Vol. 53, No. 2 (2002) 171–193

State-of-the-Art Report

Journal impact measures in bibliometric research

WOLFGANG GLÄNZEL,[1,2] HENK F. MOED[3]

[1] *Bibliometrics Service, Library of the Hungarian Academy of Sciences, Budapest (Hungary)*
[2] *Research Association for Science Communication and Information e. V. (RASCI), Berlin (Germany)*
[3] *CWTS, Leiden University, Leiden (The Netherlands)*

The Impact Factor introduced by Eugene Garfield is a fundamental citation-based measure for significance and performance of scientific journals. It is perhaps the most popular bibliometric product used in bibliometrics itself, as well as outside the scientific community. First, a concise review of the background and history of the ISI impact factor and the basic ideas underlying it are given. A cross-citation matrix is used to visualise the construction of the Impact Factor and several related journal citation measures. Both strengths and flaws of the impact factor are discussed. Several attempts made by different authors to introduce more sophisticated journal citation measures and the reasons why many indicators aiming at a correction of methodological limitations of the Impact Factor were not successful are described.

The next section is devoted to the analysis of basic technical and methodological aspects. In this context, the most important sources of possible biases and distortions for calculation and use of journal citation measures are studied. Thereafter, main characteristics of application contexts are summarised.

The last section is concerned with questions of statistical reliability of journal citation measures. It is shown that in contrast to a common misbelief statistical methods can be applied to discrete 'skewed' distributions, and that the statistical reliability of these statistics can be used as a basis for application of journal impact measures in comparative analyses. Finally, the question of sufficiency or insufficiency of a single, howsoever complex measure for characterising the citation impact of scientific journals is discussed.

Introduction

Journal citation measures are designed to assess significance and performance of individual journals, their role and position in the international formal communication network, their quality or prestige as perceived by scholars. Scientific journals may differ with respect to their importance of their position in the journal communication system, their status or prestige.

0138–9130/2002/US $ 15.00

In their review of journal citation measures published in the *Journal of Information Science* in 1987, *Todorov* and *Glänzel* characterised these measures as follows. "Journal citation indicators are commonly used as general measures for various journal characteristics and research impact by different participants in the publication, dissemination, and evaluation process of scientific knowledge. … Many librarians, information scientists and, sociologists of science already consider journal citation analysis as a practical alternative to subjective judgement. Authors may take citation measures from JCR and use them as possible indicators of journal characteristics. Lists of ranked SCI journals may help potential and real users to identify sources with significant contributions. Editors and publishers may relate high citation impact to a successful editorial practice and policy. That is why we are trying in this paper to review and comment on some citation-based measures for scientific journals which are available and applied as evaluative indicators."

The Impact Factor introduced by Eugene Garfield and regularly published in the annual updates of the *Journal Citation Reports* (JCR) is a fundamental citation-based measure for significance and performance of scientific journals. Wherever the JCR is available, a variety of journal citation measures including and beyond the Impact Factor can readily be built and calculated. If this is done based on a sound and validated methodology and if all procedures and results are properly documented, powerful bibliometric tools with wide fields of applications in the field of science information, library science, information retrieval and research evaluation can be obtained from databases originally designed for the retrieval of scientific information. Thus, especially the Impact Factor has become perhaps the most popular bibliometric product used in bibliometrics itself but also outside the scientific community. However, this popularity involves also dangers. The arbitrary treatment of data, the arbitrary weighting of components and indicators and the uninformed and tendentious use and application of journal citation measures has already done harm to the credibility of bibliometric research and technology. In fact, the use of impact factors ranges from well-documented and methodically sound applications to rather 'grey' applications as background information for scientific journalism or in the context of refereeing procedures. Impact factors are sometimes used even as substitutes for missing citation data.

The strengths of the Impact Factor lies first of all in the comprehensibility, stability and seeming reproducibility, on the other hand, some obvious flaws, but especially the already mentioned uninformed use have provoked critical and controversial discussions about its correctness and use. In this context, it has also to be mentioned that ISI's somewhat poor background documentation concerning the processing of the data presented in the JCR cannot convince critical users. In particular, the IF and related journal impact measures can readily be reproduced from the data presented in the JCR, however, these very data proved at large not to be reproducible. Although it is difficult to theoretically define the concept of (journal) impact, there is a wide spread belief that the ISI Impact Factor is affected or ‘disturbed’ by factors that have nothing to do with (journal) impact. Consequently, several attempts have been made to improve the impact factor or to develop additional or alternative journal citation measures. Some of the main modifications relate to all of the ‘elements’ mentioned in the above-mentioned mathematical interpretation.

– Instead of the mean: other parameters of the distribution (e.g. percentage of uncited papers or quantiles);
– Instead of integer counting of citations: weighting a citation on the basis of the journals in which it is made;
– Instead of applying a single citing year: application of a range of citing years;
– Instead of analysing all (‘citable’) documents: disaggregate articles on the basis of document type (article + note + review) or content (e.g., theoretical, methodological and experimental);
– Instead of considering only papers 1-2 years earlier: analysing articles from older ‘ages’;
– Instead of synchronic: diachronic, or a combination of the two approaches.

In the last two decades, several bibliometric research centres therefore succeeded in calculating their own journal impact measures on the basis of the bibliographic databases of the ISI.

In the following, we intend to give an overview of the definition and interpretations, the history, the application context and the statistical background of the impact factor, as well as of the most important attempts to improve or to complement this measure. In this context, we will focus on applications and not on technical-methodological issues, and will approach the problem of validity and usefulness from the point of view of its actual use and the context of application, that is, of the evaluative or policy questions addressed in the application.

Historical remarks

Without any doubt, the Impact Factor introduced by Garfield is the most prominent journal citation measure, and should be mentioned first. According to the *Journal Citation Reports*, the impact factor “is basically a ratio between citations and citable items published. Thus, the 1980 impact factor of journal X would be calculated by dividing the number of all the SCI source journals' 1980 citations of articles journal

X published in 1978 and 1979 by the total number of [citable] source items it published in 1978 and 1979." Thus, the impact factor is "a measure of the frequency with which the average cited article in a journal has been cited in a particular year" (*Garfield*, 1972).

The Immediacy Index is another specific journal citation indicator published regularly in the *Journal Citation Reports*. The immediacy index is "a measure of how quickly the average cited article, in a particular journal is cited. A journal's immediacy index considers citations made during the year in which the cited items were published" (cf. *Journal Citation Reports*). This indicator is even more affected by a number of technical conditions such as the publication delay, frequency of publication, speed of indexing, subject peculiarities (ageing) and the document type than the impact factor itself. These conditions resulted in limitations concerning the importance and the use of this indicator.

Because of its comprehensibility, robustness and its fast availability, the impact factor became very quickly popular and widely used. The Impact Factor is *comprehensible* because it measures the frequency with which an average article published in a given journal has been cited in a particular year; it is *robust* because the annual changes of the journals' Impact Factors proved to be not dramatic so that in practice one or two years old impact factors are sometimes used for evaluation purposes where more recent indicators are not available. On the other hand, time series can be used to monitor the evolution of journals' citation patterns. The *fast availability* of the Impact Factor, finally, is due to the fast indexing, data processing and the distribution of ISI products. These are in short the most important technical advantages of the Journal Impact Factor.

On the other hand, according to a number of authors both the Impact Factor and especially the Immediacy Impact have several serious flaws the consequences of which shall be discussed here.

1. There is no normalisation for reference practices and traditions in the different fields and disciplines (*Pinski* and *Narin*, 1976).
2. "There is no distinction in regard to the nature and merits of the citing journals" (*Tomer*, 1986).
3. There is a bias in favour of journals with lengthy papers, e.g. review journals tend to have higher impact factors (*Pinski* and *Narin*, 1976).
4. Citation frequency is subject to age bias (*Asai*, 1981; *Rousseau*, 1988; *Glänzel* and *Schoepflin*, 1995; *Moed* et al., 1998).
5. There is no indication of the deviations from this statistic (see, for instance, *Schubert* and *Glänzel*, 1983).

6. The average time for a journal article from publication to peak in citations is not always two years, or as *Garfield* (1986) writes "if we change the two-year based period used to calculate impact, some type of journals are found to have higher impacts" (cf. also *Glänzel* and *Schoepflin*, 1995, *Moed* et al., 1998).
7. One single measure might not be sufficient to describe citation patterns of scientific journals.
8. The concept of citable document is not operationalised adequately. As a result, journal impact factors published in ISI's *Journal Citation Reports* are inaccurate for a number of journals (*Moed* and *van Leeuwen*, 1995, 1996).
9. In the calculation of JCR impact factors, errors are made due to incorrect identification of (cited) journals, for instance for the journal *Angewandte Chemie – International Edition* (*Braun* and *Glänzel*, 1995, *van Leeuwen* et al., 1997).

The above-mentioned limitations lead very early to the discussion of possible improvements or alternatives. Among others, *Yanovski* (1981) criticised certain distortions of the impact factor and suggested new indicators based on the ratio between citations and references. *Yanovski*'s indicator has not found wider use although, for instance, *Smart* and *Elton* (1982) and *Todorov* (1983) have critically reacted on his approach. Thus, *Smart* and *Elton* showed that the consumption factor and the impact factor are statistically independent which suggests that these two measures represent distinct journal attributes.

In 1978 *Lindsey* introduced the Corrected Quality Ratio defined as (number of citations)$^{3/2}$/(number of publications)$^{1/2}$. This formula can be reformulated as the product of the square root of the impact factor and the number of citations. This approach, however, lacks interpretability. This might be one reason why this indicator has not found application.

An interesting and undeservedly neglected approach has been given by *Allison* in 1980. He used the statistical function (standard deviation – mean value)/(mean value)^2 as an inequality measure of distributions of scientific productivity and citation impact. The underlying assumption is that the distribution of authors by the number of publications or by the frequency of citations is negative binomial. This is one of the first times that a particular distribution model is assumed for citation frequency. The indicator is the reciprocal of an estimator of one of the parameters N of the distribution.

Schubert and *Glänzel* (1983) have studied the statistical reliability of journal Impact Factors. In particular, they analysed both the significance of the deviation between the impact factors of the same journals calculated by different institutes and that of the deviation between the impact factors of two different journals representing the same discipline. They have used a similar model as suggested by *Allison*. The results of this study have strong methodological and practical influence on journal rankings by impact factors and comparative analyses in research evaluation.

According to *Asai* (1981) the period count based on a month produces more accurate statistics than that based on a year. The author introduces an Adjusted Impact Factor which counts the weighted sum of citations over a period of four years instead of one year as in case of the original Impact Factor.

The most sophisticated improvement has been presented by *Narin* and *Pinski*. Whereas in calculating the impact factor and the immediacy index all citations are equally weighted, the "influence methodology" suggested by *Pinski* and *Narin* (1976) provides for each journal a size-indepedent Influence Weight determined by the number of the journal's citations and references. The calculation is based on an iteration procedure involving great expense. Weighting citations by these influence weights, the *influence per publications* and the *total influence* can be calculated. *Geller* (1978) suggests a 'corrected' influence weight that could be interpreted as the probability that a given journal will be cited from the other journals. Because of the troublesome calculations and the lack of expressive interpretability of the results, the method has gained few adherents.

Since one single impact measure might not be sufficient to describe citation characteristics of journals, supplementary indicators have been introduced. The most simple, robust, readily interpretable and reproducible indicator of this type is the share of uncited papers or cited papers, respectively (cf., *Schubert* and *Glänzel*, 1983; *Moed* et al., 1999).

The relationship between journal citation indicators can best be shown by a journal cross-citation matrix (cf. Table 1). In the already mentioned review paper by *Todorov* and *Glänzel* (1987), this is used to visualise the construction of the Impact Factor and several related journal citation measures.

The traditional citation transaction matrix consists of journal titles in the rows (citing sources) and the columns (cited journals), and is formed by 'cells' C_{ij}. Each cell records the number of references C_{ij} that the source journal J_{si} gives to a destination (cited) journal J_{dj}. The row sum $R_i = C_{i1} + \ldots + C_{im}$ represents the total number of references given from J_i to other tabulated journals. The element C_j indicates also the number of citations received by J_{dj} from J_{sj}. The column sum $C_j = C_{1j} + \ldots + C_{nj}$ is the total number of citations received by J_{dj} from the matrix source journals. This important type of matrix is representing the process of transaction of references (or citations) among journals. It is asymmetric with relatively high values in the diagonal where self-citations are registered. Usually, only citable papers (e.g., articles, letters, notes, reviews) of J_{si} and J_{dj} published in the periods t_r and t_c, respectively, are taken into consideration. Finally let $|J_{**}|$ denote the number of citable papers published in the journal J_{**} in the period t_* (t_* usually denotes the period of one year).

Table 1. Cross-citation matrix representing transaction of references (citations) among source (citing) and cited journals

		Period t_c					
		J_{d1}	…	J_{dj}	…	J_{dm}	Σ
Period t_r							
	J_{1s}	C_{11}	…	C_{1j}	…	C_{1m}	R_1
	…	…		…		…	…
	J_{is}	C_{i1}	…	C_{ij}	…	C_{im}	R_i
	…	…		…		…	…
	J_{ns}	C_{n1}	…	C_{nj}	…	C_{nm}	R_n
	Σ	C_1	…	C_j	…	C_m	–

A number of journal citation indicators can be defined using these notations. In Table 2, we give the definitions according to the above table for the Impact Factor, Immediacy Index, Adjusted Impact Factor, Consumption Factor and Influence Weight just as examples. Formulae for further journal citation measures can be found in *Todorov* and *Glänzel* (1987).

Table 2. Selected journal citation measures defined on the basis of a cross-citation transaction matrix

Journal impact measure	Definition
Impact Factor	$\mathrm{IF} = C_j(t_c)/(\lvert J_d(t_{c-2})\rvert + \lvert J_d(t_{c-1})\rvert)$
Immediacy Index	$\mathrm{II} = C_i(t_c)/\lvert J_d(t_c)\rvert$
Consumption Factor	$\mathrm{CF} = C_i(t_c)/R_i(t_c)$
Adjusted Impact Factor	$\mathrm{AIF} = (0.250\cdot C_i(t_c) + 0.875\cdot C_i(t_{r+1}) + 0.875\cdot C_i(t_{r+2}) + 0.875\cdot C_i(t_{r+3}))/\lvert J_d(t_r)\rvert$
Influence Weight	$\mathrm{w}_i = \Sigma_k\, w_k\, C_{ki}/R_i$

Most journal citation indicators aiming at a correction of some of the methodological limitations of the Impact Factor were not truly successful. The reasons for this phenomenon can be summarised as follows.

Apart from the impact factor, the use of other journal citation measures is rather occasional. In spite of its availability, the immediacy index has because of the already mentioned technical and methodological shortcomings only limited significance. All other attempts to improve, substitute or supplement the impact factor have encountered serious obstacles to wider application. On the one hand, the achieved 'accuracy' is often at the expense of simple interpretation. On the other hand, several alternative journal citation measures could not always be rendered accessible to a broader audience, or at least not regularly be updated like in the case of the IF through the annual updates of the JCR. In lack of regular updates, these indicators could not be readily reproduced by other research centres.

Nowadays, improvements therefore focus more on the particular choice of publication period and citation window, the calculation of separate indicators for different document types, the development of 'relative', field-normalised measures, and the use of supplementary measures and the clarification of the technical correctness of the processed indicators. These attempts will be discussed in the following section.

In addition to the above-mentioned citation measures, we also refer to several time related indicators of journal citation such as the Price Index (*Price*, 1970), (Citing and Cited) Half Life (JCR, annually since 1975; *Moed* et all, 1998) and Mean Response Time (*Schubert* and *Glänzel*, 1986). These indicators are designed to measure ageing characteristics of subject fields and journal literature, or to help to distinguish between 'hard' and 'soft' sciences or between slow and fast reception of scientific information. Although these measures do not reflect any aspects of 'impact', 'status' or 'prestige' of a journal, they have to be at least mentioned here since they are by definition true journal citation indicators, too.

Technical and methodological aspects

In this section, elementary technical and methodological aspects will be analysed in the light of practical use, of reproducibility, comprehensibility and interpretation of journal impact measures.

Journal impact measures have – as all citation-based measures – field-specific biases. In particular, citation impact is mainly influenced by at least the following five factors:

1. the document type;
2. the subject matter;
3. the paper's age;
4. the paper's 'social status' (through the author(s), the author's institution and the journal);
5. the observation period ('citation window').

These factors are heavily influencing impact measures, and are thus characteristic for citation peculiarities of publications and journals, but do not belong to the above-mentioned 'disturbing factors'.

Before discussing these factors and approaches, we first have to point to a basic accuracy problem in determining citation rates to particular journals. In calculating the impact factor of a journal, ISI attempts to identify all citations to that journal, on the basis of the information on the cited source (journal title) in all cited references processed. It should be noted that identifying a particular journal in a collection of millions of cited journal strings is not always an easy task. Important variations of a journal title may be overlooked. In addition, a cited journal string may not provide sufficient information for establishing which journal was actually cited. Inaccuracies of this type are almost inevitable when cited references are matched to journals merely on the basis of the journal title included in a cited reference. However, they may considerably affect the impact measurement of particular journals. As a case study, *Braun* and *Glänzel* (1995), later followed by *van Leeuwen* et al. (1997) analysed the accuracy of the JCR impact factor of *Angewandte Chemie – International Edition.*

The following example shows the complexity of influences and biases in calculating impact measures. The citation mean of papers published in the two journals *American Sociological Review* (*ASR*) and *The Lancet* are compared in dependence of time. The publication year is 1980, the citation window ranges from 1, 2, 3 and 6 to 10 years. Besides subject peculiarities and the different ageing the document type is responsible for the deviating trends of the journals. Note that a considerable part (more than 60%) of all documents published in Lancet in 1980 were *letters*. The mean citation rates for both journals are presented in Table 3. In citation windows shorter than 4 years, the citation impact of *The Lancet* exceeds that of the *American Sociological Review*. Four years after publication, the social sciences journal exhibits a clearly greater impact than the life science journal.

Next, the question of citable documents is discussed. Not all document types proved conveyers of relevant scientific information. Within relevant document types, there is a certain bias in favour of *reviews* and to the detriment of *letters* (*Braun* et al., 1989, *Moed* and *van Leeuwen,* 1995). Table 4 visualises the mean citation rates of different document types in selected journals. (Publication year: 1995 + 1996, Citation window: 1995-1997 + 1996-1998.)

Table 3. Mean citation rate of two journals in time as a function of time
(source year: 1980)

Citation window	Mean Citation Rate *ASR*	Mean Citation Rate *Lancet*
1980-80	0.2	0.6
1980-81	1.8	2.4
1980-82	4.3	4.5
1980-85	12.1	9.7
1980-89	20.9	14.0

Table 4. "3-year impact measure" for selected journals by document types

#	Journal	Mean Citation Rate			
		Total	Articles	Reviews	Letters
1	*Science*	32.86	42.30	145.35	0.41
2	*Nature*	32.88	49.73	96.07	3.93
3	*Lancet*	5.25	17.55	14.68	1.99
4	*Cell*	75.68	74.82	78.63	75.64
5	*Angew. Chem. Int. Ed.*	11.01	9.37	32.03	19.00
6	*J. Acq. Immun. Defic. Synd. Hum. R.*	4.05	4.64	39.00	1.04
7	*Int. J. Rad. Oncol. Biol. Phy.*	3.52	4.15	35.00	0.37
8	*J. Phys. Condens. Matter*	2.72	2.47	9.57	3.99

The data presented in Table 4 also show that at least two types of letters have to be distinguished. *Moed* and *van Leeuwen* (1995) argue in this context: "Our results suggest that differences exist among subfields of science. In clinical medicine, case reports or letters are rather short communications, typically one page long, and have on average a lower impact than normal articles, while in physics or astronomy letters are somewhat longer, are more similar to normal articles- also with respect to their impact – but are generally published more rapidly". Thus letters published in *Science, Nature, Lancet, Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, International Journal of Radiation Oncology Biology Physics* exhibit a significantly lower impact than articles or reviews of the same journals. Other types of letters might have even higher citation impact than articles (cf., *Cell*, *Angewandte Chemie – International Edition*, *Journal of Physics – Condensed Matter*). On the other hand, *reviews* receive, on the average, the highest citation rates among these three document types. All other types, for instance, book reviews, discussion papers, editorial material or meeting abstracts proved not to contain original research results and/or are not likely to be cited.

As indicated above, the IF of a journal in year *T* is defined as the number of citations in year *T* to documents published in that journal in years *T*-1 and *T*-2, divided by the number of citable documents published in that journal in years *T*-1 and *T*-2 (*Garfield*, 1979). However, the concept of citable document is not defined accurately by ISI. As pointed out by *Garfield* for 40 leading medical periodicals, journals differ with respect to the numbers and types of documents they publish, and variations exist in impact for different types (*Garfield*, 1987). *Moed* and *van Leeuwen* (1995, 1996) obtained evidence that the IF of many journals included in the *Science Citation Index* (SCI) are inaccurate, due to an inappropriate definition of citable documents. In particular, ISI classifies documents into types. In calculating the numerator of the IF, ISI counts citations to all types of documents, whereas as citable documents in the denominator ISI includes as a standard only normal articles, notes and reviews. However, editorials, letters, and several other types are cited rather frequently in a number of journals. When they are cited, these types do contribute to the citation counts in the IF's numerator, but are not included in the denominator. In a sense, the citations to these documents are 'for free'. This phenomenon is illustrated in Table 5 for one particular journal: the *Lancet*. The citation data relate to the citing year 1992, and the publication data to the years 1990 and 1991. The Table shows that the *Lancet* has published large numbers of letters and editorials. In the 'correct' approach, the journal's IF is calculated as the number of citations to articles, notes and reviews (b), divided by the number of articles, notes and reviews (a). This ratio amounts to 8.3. According to the ISI JCR approach, however, the IF is defined as the total number of citations to all types (c), divided by the number of articles, notes and reviews (a), and amounts to 14.7. Consequently, the 'correct' impact factor of this journal in 1992 would be 43 percent lower than the IF listed in the JCR.

Citation patterns are strongly influenced by subject characteristics. Citation measures in general and impact factors in particular are therefore – without normalisation – not appropriate for cross-field comparisons. The following example visualises the subject bias of citation measures (see Table 6).

Several attempts have been made to compensate citation-specific biases such as the subfield characteristics. For instance, the Disciplinary Impact Factor (DIF) was introduced by *Hirst* (1978) as the average number of times a journal is cited in a given subfield alone rather than across the complete set of the SCI. *Van Leeuwen* and *Moed* developed a journal-to-field-impact score (*Moed* et al, 1998, *van Leeuwen* and *Moed*, this issue).

Table 5. ISI impact factors are inaccurate: *The Lancet* 1992 (from *Moed* et al., 1999)

Type of article	*P*	*C*	*C/P*
Articles	784	7134	9.1
Notes	144	593	4.1
Reviews	29	232	8.0
SUBTOTAL	957 (a)	7959 (b)	8.3
Letters	4181	4264	1.0
Editorials	1313	905	0.7
Other	1421	909	0.6
TOTAL	7872	14037 (c)	1.8
ISI JCR Impact factor:	(c) / (a) = 14.7		
Correct Impact factor:	(b) / (a) = 8.3		

Table 6. Mean citation rate of subfields
(source year: 1996, citation window: 1996-1998)

Mechanical, civil and other engineering	1.12
Mathematics	1.46
Analytical chemistry	3.00
Solid state physics	3.06
Neurosciences	4.54

Schubert et al. (1987) and *Glänzel* and *Schubert* (1988) have chosen a different way by developing a method which offers an optionally refinable and self-adjusting scale with variable scores. Although the scores do not depend on any particular citation distribution, these scores have interesting properties if the underlying citation distribution is Paretian. The scores are defined in a recursive manner. 0 is chosen as the first score b_0. The second one (b_1) is the mean citation rate of the journal, the third score (b_2) is the mean citation rate of those papers published in the journal in question which have received more citation than the previous score, and so on. The scores proved to characterise the underlying citation distribution, and were therefore called ‘characteristic scores’. The method itself was called ‘characteristic scaling’. The scores themselves are, of course, reflecting subject peculiarities but the series of shares of papers having received 0 citations, more than b_i but less than b_{i+1} ($i = 0, 1, 2 \ldots$) citations is rather expressing journal characteristics than subject-specific properties (cf., *Schubert* et al., 1989).

The notion of ageing or obsolescence is closely related with the age structure of citations or references (‘diachronous’ and ‘synchronous’ approach, respectively). First *Wallace* (1986) studied the relationship between journal productivity and obsolescence.

He assumed an exponential distribution, consequently, the ageing can be characterised by the 'half-life' being the median of the distribution.

According to the diachronous model, the ageing of scientific literature is reflected by the 'citation life-time curve', that is, the distribution of citations to a given set of papers over time.

The ageing process can be considered a composition of two overlapping time intervals, one reflecting the *maturing phase*, and the other one referring to the phase of *decline*. If in particular, the maturing phase corresponds to an initial period of progressive growth described by growing increments, the decline phase has, on the contrary, to be considered as a period of saturation which is characterised by diminishing increments. *Glänzel* and *Schoepflin* (1995) and *Moed* et al. (1998) have distinguished four major types of ageing processes according to different periods of maturing and decline. They have shown that these types are characterising life-time curves of journals. The four types are described in Table 7.

Table 7. Four different types of ageing processes

Ageing type	Period of maturing	Period of decline
Type I	short	short
Type II	long	short
Type III	short	long
Type IV	long	long

The following two examples (cf. Figs 1 and 2) show the life-time curves of two journals of different types in 1980. The journal *Proceedings of the National Academy of Sciences of the United States of America* represents Type I, whereas the ageing curve of the journal *Annals of Statistics* corresponds to the Type IV. Both journals represent almost extreme cases. In the first example, the peak of the citation curve is reached in 1982, that is, two years after publication. In the second case, the peak is reach only 5 years after publication, namely in 1985. *PNAS US* received 31.2% of all citations observed till 1995 in the initial period 1980-1982, *Annals of Statistics* only half as much (16.5%). The different ageing behaviour has strong influence on possible selection of the citation window underlying the construction of the citation indicator.

Moed et al. (1998) have proposed a new classification system of journals in terms of their ageing characteristics is introduced. This system has been applied to as many as 3,098 journals covered by the *Science Citation Index* (SCI). Following an earlier suggestion by *Glänzel* and Schoepflin (1995), a maturing and a decline phase were distinguished.

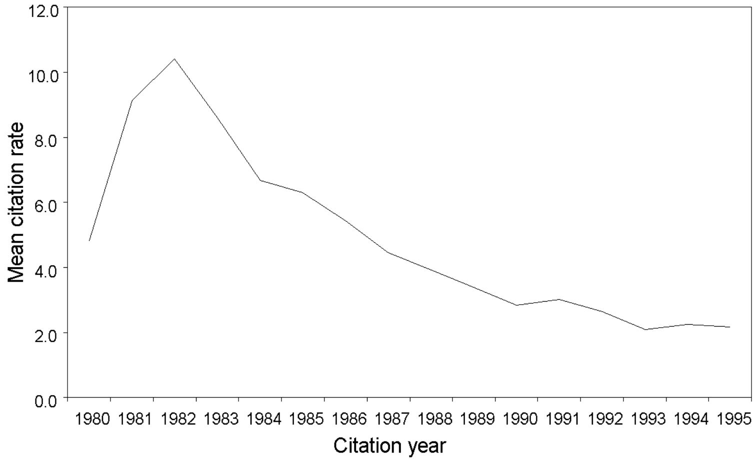

Figure 1. Citation life-time curve of papers published in the journal *PNAS US* in 1980

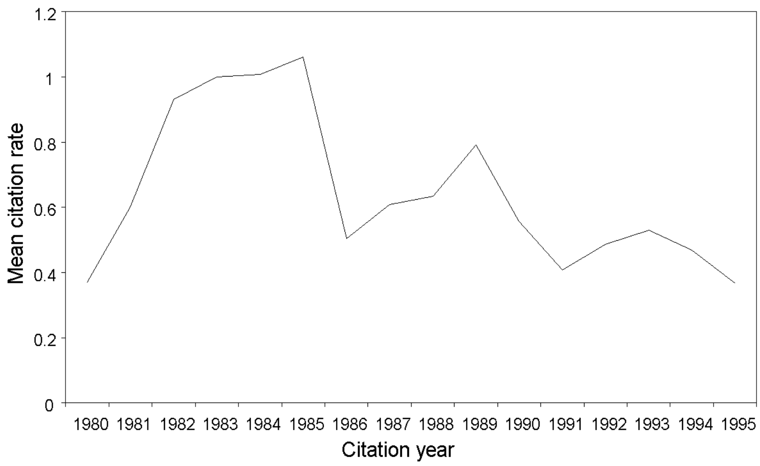

Figure 2. Citation life-time curve of papers published in the journal *Annals of Statistics* in 1980

From an analysis across all subfields the authors concluded that ageing characteristics are primarily specific to the *individual* journal rather than to the *subfield*, while the distribution of journals in terms of slowly or rapidly maturing or declining types is specific to the subfield. The current JCR journal impact factor is shown to be biased towards journals revealing a rapid maturing and decline in impact.

The authors also found that the Cited Half Life (CHL), printed in the JCR, is an inappropriate measure of decline of journal impact. A more adequate parameter of decline is proposed, taking into account the size of annual publication volumes during a range of 15 years. It was found that for 76% of SCI journals the relative difference between this new parameter and the ISI CHL exceeds 5 percent.

Moed et al. (1998) proposed a longer term impact factor, as well as a normalised impact statistic, taking into account both citation characteristics of the research subfield covered by a journal as well as the type of documents published in it. The authors argue that "when these new measures are combined with the proposed ageing classification system, they provide a significantly improved picture of a journal's impact than that obtained from the JCR".

Other limitations result from the particular choice of the citation window. In this context, the 'diachronous' versus 'synchronous' approach is discussed. Recent bibliometric studies are increasingly using 'diachronous' impact factors based on citation windows larger than one year. In recent papers, *Ingwersen* et al. (2000, 2001) have given a methodological discussion why the 'diachronous' approach should be preferred to the 'synchronous' one.

From the viewpoint of practical experience, we can conclude the following. At the LHAS in Hungary and at RASCI in Germany citations are often counted in a three-year observation period, particularly, in the year of publication and the two subsequent years. This practice has been applied since 1995. According to results by *Glänzel* and *Schoepflin* (1995), the 3-year citation window proved to be a good compromise between the relatively fast obsolescence of technology oriented literature, of most areas in life sciences, of experimental physics literature, on one hand, and of the slowly ageing theoretical and mathematical topics in physics, on the other hand.

The above-mentioned attempts to overcome technical and methodological shortcomings of the impact factor result, on the one hand, in methodological improvements but they cause, on the other hand, also limitations to the reproducibility of the published outcomes for users.

Application contexts

One of the crucial questions in the application context has already been mentioned in the last section. The universal availability of the Impact Factor (together with its comprehensibility) forms the main strength of this indicator. Any modification of its original 'definition' necessarily results in an incompatibility to the ISI Impact Factors used by thousands of scientists in a professional or semi-professional way. Normally, these modified measures cannot be reproduced with the help of journal data published in the *Journal Citation Reports.* But even measures based on data from the JCR may lack reliability, as the data published in this volume may be inaccurate.

In the context of possible application, one has to distinguish between:

a) *Different purposes*
From the viewpoint of purposes, the application to scientific information, the use of journal impact measures in library and information management and the use in research evaluation and science policy should be mentioned.

b) *Different levels of aggregation*
Among the levels of aggregation, three types have to be distinguished:
(i) the micro level (individual scientist, research group),
(ii) the meso level (institutions, journals) and
(iii) the macro level (national and supra-national research, subject analyses).

The question, which journal indicators are the most appropriate, might have different answers depending on the field of application. The fundamental question, which aspects of journal performance these measures are assumed to indicate and how particular quantitative measures of journal impact correlate with the perceptions of scientists on 'significance', importance or prestige of journals in a particular field, is, however, of universal validity for each purpose and at all levels of aggregation.

Statistical reliability of journal citation measures

The previous sections were, among others, concerned with the bibliometric/ /methodological interpretation of the Impact Factor and related journal citation measures. It has to be mentioned that these measures have also a mathematical interpretation beyond that introduced through the journal cross-citation matrix. Most journal citation indicators can be interpreted as statistical functions, and can therefore be analysed with the tools of mathematical statistics. In particular, Immediacy Index, Impact Factor and all their modifications are mean values of citation distributions over journal publications and the share of (un-)cited papers can be interpreted as an estimate of the corresponding probability of (un-)citedness.

In contrast to the common misbelief statistical methods can be applied to discrete 'skewed' distributions and the statistical reliability of these statistics can be used as a basis for application of journal impact measures in comparative analyses. Under the condition that the discrete distribution in question belongs to the domain of attraction of the normal distribution, the sample mean, say $\overline{x}$, has the distribution $N(m, \sigma)$ where m is the expectation of the underlying (discrete) distribution and σ depends only on the standard deviation of the underlying distribution and the sample size. The above condition implies that the law of large numbers holds and we have $E(\overline{x}) = m$ and $D(\overline{x}) = \sigma = d/\sqrt{n}$, where d is the standard deviation of the underlying (discrete) distribution and n the sample size. Moreover, $\overline{x}$ is an unbiased estimator of m and the standard deviation of the sample mean is an increasing function of the distribution's standard deviation and a decreasing function of the sample size. Many discrete distributions satisfy the above condition. Among these distributions, the negative binomial, the geometric, the Poisson distribution, the log series distribution, all finite distributions such as the binomial and hyper-geometric distribution and all Paretian distributions with exponent > 2 can be found. These properties can be used as a basis for comparisons with the impact of other journals and allows to decide whether the deviation of the impact of a journal from that of others is significant or not (see *Schubert* and *Glänzel*, 1983). The following examples illustrate how these techniques work, and that any evaluative conclusions drawn from linear ranking of journal citation measures might be questionable.

Examples

The first example has been taken from *Schubert* and *Glänzel* (1983). We compare the 'corrected impact factor' x_1 calculated at ISSRU (Budapest) on the basis on articles, letters notes and reviews with the JCR impact factor x of the journal *Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie* (papers published in 1978/79, citations counted in 1980). A simple Welch test is applied. According to this test, the deviation of two Gaussian random variables is considered significant if the absolute value of ratio of the difference of the two variables and their standard deviation exceeds the critical value belonging to the given confidence level. If one of the two variables is a constant, that is, a given fixed value, the standard deviation of the remaining random variable appear in the denominator. We will consider the JCR impact factor a given fixed value.

ISSRU: $x^{*} = 0.611$ $D(x^{*}) = 0.078$
JCR: $x = 0.532$

Thus, we have

$$w = \frac{x^* - x}{D(x^*)} = \frac{0.611 - 0.532}{0.078} = 1.01 \cdot$$

Since $|w| = 1.01 < 1.96 = w_p$ ($p = 0.95$), the deviation is not significant at a confidence level of 0.95. In verbal terms, the deviation of the 'corrected impact factor' from the JCR Impact Factor of this journal in 1980 is not significant although the ISSRU value exceeds the corresponding JCR value by almost 15%. Such statement does, however, not hold for all journals, as the second example will show.

In the following, we compare the ISSRU value of *Angewandte Chemie – International Edition* with the corresponding JCR Impact Factor (papers published in 1978/79, citations counted in 1980).

ISSRU: $x^* = 3.300$ $D(x^*) = 0.117$
JCR: $x = 4.769$

Now, we have

$$w = \frac{x^* - x}{D(x^*)} = \frac{3.300 - 4.769}{0.117} = -12.56 \cdot$$

Since $|w| = 12.56 > 1.96 = w_p$ ($p = 0.95$), the deviation has to be considered significant at any reasonable confidence level. The reason for the large deviation between these two impact measures has already been discussed in the section concerned with technical and methodological aspects.

According to Glivenko's theorem, the empirical distribution converges to the underlying theoretical one with probability 1. The relative frequency f is an unbiased estimator of the corresponding probability p. For instance, the standard error of the share of cited papers (f_c) can be calculated analogously to the impact factor. In particular, we have $E(f_c) = p_c$, where p_c is the corresponding probability that a paper will be cited and $D(f_c) = \sqrt{p_c\,(1 - p_c)/n}$ .The standard deviation of the relative frequency is a decreasing function of the sample size. Concerning the dependence on p_c, there is a global maximum for $p_c = ½$**.**

The next question refers to the sufficiency or insufficiency of a single, howsoever complex measure for characterising the citation impact of scientific journals. This question arises in the mirror of the analysis of the shape of journal citation distributions. Moreover, the shape of such citation distributions might change over time, for instance, from extremely skewed to less skewed distributions with peak at citation rates greater than 0. Figure 3 presents the distribution of citation over articles and reviews published in the journals *Angewandte Chemie – International Edition* (*ACH*) and *JACS* in

1995/96. Citations have been counted in the period 1995-1997 for papers published in 1995 and in 1996-1998 for papers published in 1996. The corresponding Mean Citation Rates are with 11.06 for *ACH* and 11.26 for *JACS* almost identical. Nevertheless, the shapes of the two distributions are quite different. While uncited paper are rare in *JACS*, the share of uncited papers in *ACH* amounts to almost 10%. This is compensated by a relatively high share of highly cited papers in *ACH*. However, it should be mentioned in passing that the large number of citations is mainly attracted by reviews. In 1995/96, the share of reviews in all citable papers published in *ACH* (7.5%) was twice as high than the corresponding share of *JACS* (3.6%). On the other hand, if reviews are not taken into account, that is, if only articles are considered, the share of 'highly cited' papers with at least 50 citations each of *JACS* (10.4%) exceeds that of *ACH* (7.5%) by far. The share of uncited papers, however, remains practically unchanged.

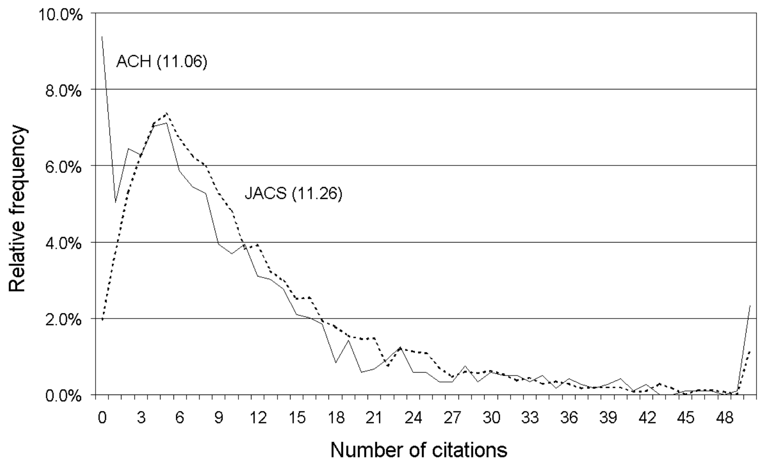

Figure 3. Example for citation distributions: *JACS* and *Angewandte Chemie – International Edition (ACH)*

The time-dependence of citation impact can be modelled with the help of stochastic processes. In this context, mathematical tools might help to find optimum solutions for the choice of citation windows and indicator sets for measuring journal citation impact.

In order to reflect the influence of the above-mentioned factors (subject matter, ageing, social status, document type, citation window, non-homogeneity), the distribution model for citation impact should have at least two free parameters, one of which has to be time-dependent. *Glänzel* and *Schubert* (1995) and *Glänzel* and

Schoepflin (1994, 1995) introduced a negative-binomial process (a special case of a non-homogeneous birth-process) as a model for the change of citation impact in time and for the ageing of scientific literature. The non-homogeneity of the process is visualised by the following example presented in Table 8.

Figure 4 shows the changing shape of citation distributions. The trend from an extremely skewed towards a less skewed distribution is obvious. The (1-year) impact measure of 0.14 increases to 5.60 for the 16-year period 1980-1995. A very convenient solution to stay in keeping with the non-homogeneity of the stochastic process is to find a supplementary indicator (preferably a measure which is independent of time since the time-dependence is already expressed by the impact measure).

Table 8. Mean observed citation rates of papers published in 1980 based on 3-year citation windows in an initial period and ten years later

Journal	1981-1983	1991-1993
Tetrahedron	5.58	3.55
Blood	14.2	5.77
Physical Review D	8.27	3.37
Nature	20.81	8.93

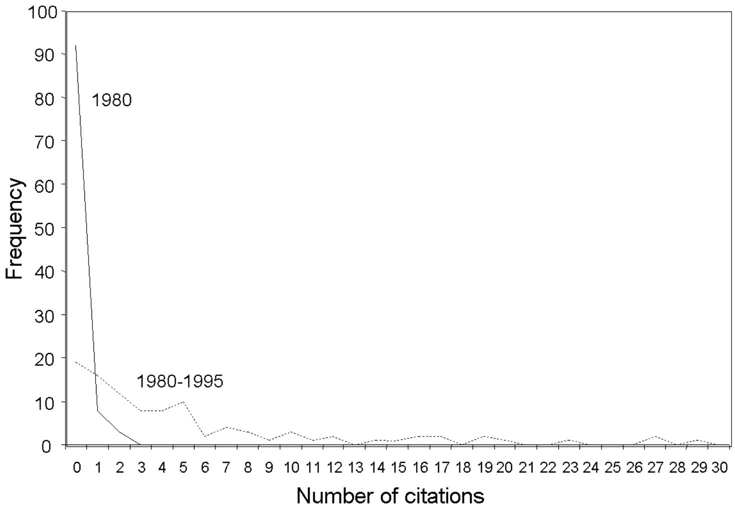

Figure 4. Changing shape of the journal *Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*

One solution has already been mentioned. The share of (un)-cited papers (as a supplementary measure) and the impact factor completely describe a model of a non-homogeneous process with two free parameters. This indicator has been used by *Schubert* and *Glänzel*, (1983) and *Moed* et al. (1999) who have given empirical evidence that the two measures (impact and share of uncited papers) are statistically not independent. Moreover, the share of uncited or cited papers changes in time (see Figure 4). According to the non-homogeneous negative-binomial birth-process used by *Glänzel*, *Schubert* and *Schoepflin* the relationship between two indicators x (impact factor) and f_0 (share of uncited papers) is restricted by the inequality $\overline{x} > -\log f_0$, that is, the two measures can also theoretically not be independent. Now, we can take up the approach by *Allison* (1980). As mentioned above, he used the statistical function (standard deviation – mean value)/(mean value)^2 as an inequality measure of distributions of scientific productivity and citation impact. It has been shown that under the assumption of a negative-binomial distribution model this indicator is the reciprocal of an estimator of the parameters N of the distribution. In the model used by *Glänzel*, *Schubert* and *Schoepflin*, N is independent of time and it can be shown that the two statistical functions, $\overline{x}$ and the estimator of N are independent.

The results derived from the mathematical-statistical approach show that impact factors and related journal citation indicators should be regarded as statistical functions, and therefore be used according to the basic laws of statistics and probability theory. This applies mainly to application in comparative and evaluative studies.

Conclusions

There are new, exciting challenges in bibliometric citation analysis. The robustness, comprehensibility, methodological reproducibility, apparent simplicity, availability and popularity of the ISI journal impact factor is contrasted by several severe methodological shortcomings and its technical irreproducibility. Nowadays, it became quite tempting to apply the impact factor as a universal bibliometric measure, to use it even as a surrogate of observed citation rates. This is certainly one source of possible uninformed use.

Recent bibliometric studies have, however, shown that methodological improvements in combination with additional, multi-dimensional measures and appropriate methodological and technological documentation may help to overcome limitations of the original measures. The question of reproducibility can thus be solved at least for those who have access to the underlying bibliographic databases and the technology to reproduce these indicators. Thus, alternative journal impact measures are becoming available, ready to be diffused among wide groups of users.

References

ALLISON, P. (1980), Inequality of scientific productivity, *Social Studies of Science*, 10 : 163–179.

ASAI, I. (1981), Adjusted age distribution and its application to Impact Factor and Immediacy Index, *Journal of the American Society for Information Science*, 32 :172–174.

BRAUN, T., GLÄNZEL, W. (1995), On a source of error in computing Journal Impact Factors, *Chemical Intelligencer*, January, 31–32.

BRAUN, T., GLÄNZEL, W., SCHUBERT, A. (1989), Some data on the distribution of journal publication types in the Science Citation Index Database, *Scientometrics*, 15 : 325–330.

GARFIELD, E. (1972), Citation analysis as a tool in journal evaluation, *Science*, 178 : 471–479.

GARFIELD, E., *Citation Indexing: Its Theory and Applications in Science, Technology and Humanities,* New York: Wiley, 1979.

GARFIELD, E. (1986), The evolution of physical chemistry to chemical physics, *Current Contents*, 3 : 3–12.

GARFIELD, E. (1987), Why are the impacts of leading medical journal so similar yet so different: Item-by-item audits reveal a diversity of editorial materials, *Current Contents*, January 12.

GELLER, N. L. (1978), Citation influence methodology of Pinski and Narin, *Information Processing and Management*, 14 : 93–95.

GLÄNZEL, W., SCHUBERT, A. (1988), Characteristic scores and scales in assessing citation impact, *Journal of Information Science*, 14 : 123–127.

GLÄNZEL, W., SCHUBERT, A. (1995), Predictive aspects of a stochastic model for citation processes, *Information Processing & Management*, 31 : 69–80.

GLÄNZEL, W., SCHOEPFLIN, U. (1994), A stochastic model for the ageing analyses of scientific literature, *Scientometrics*, 30 : 49–64.

GLÄNZEL, W., SCHOEPFLIN, U. (1995), A bibliometric study on ageing and reception processes of scientific literature, *Journal of Information Science*, 21 : 37–53.

HIRST, G. (1978), Discipline Impact Factor: a method for determining core journal lists, *Journal of the American Society for Information Science*, 29 : 171–172.

INGWERSEN, P., LARSEN, B., WORMELL, I. (2000), Applying diachronous citation analysis to research program evaluations, In: CRONIN, B., ATKINS, H. B. (Eds). *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield,* Medford, N. J.: Information Today, Inc. & American Society for Information Science, pp. 373–387.

INGWERSEN, P., LARSEN, B., ROUSSEAU, R., RUSSELL., J. (2001), The publication-citation matrix and its derived quantities, *Chinese Science Bulletin*, 46 : 524–528.

LINDSEY, D. (1978), Corrected quality ratio: A composite index of scientific contribution to knowledge, *Social Studies of Science*, 8 : 349–354.

MOED, H. F., VAN LEEUWEN, TH. N. (1995), Improving the accuracy of the Institute for Scientific Information's Journal Impact Factor, *Journal of the American Society for Information Science*, 46 : 461–467.

MOED, H. F., VAN LEEUWEN, TH. N. (1996), Impact Factors can mislead, *Nature*, 381 : 186.

MOED, H. F., VAN LEEUWEN, TH. N., REEDIJK, J. (1996), A critical analysis of the Journal Impact Factors of *Angewandte Chemie* and the *Journal of the American Chemical Society*: Inaccuracies in published impact factors based on overall citations only, *Scientometrics*, 37 : 105–115.

MOED, H.F., VAN LEEUWEN, TH. N., REEDIJK, J. (1998), A new classification system to describe the ageing of scientific journals and their impact factors. *Journal of Documentation*, 54 : 387–419.

MOED, H.F., VAN LEEUWEN, TH. N., REEDIJK, J. A. (1999), Towards appropriate indicators of journal impact, *Scientometrics*, 46 : 575–589

PINSKI, G., NARIN, F. (1976), Citation influence for journal aggregates of scientific publications, *Information Processing and Management*, 12 : 297–312.

PRICE, D. J. DE SOLLA (1970), Citation measures of hard science, soft science, technology, and non-science, In: C. E. NELSON, D. K. POLLAK (Eds), *Communication among Scientists and Engineers* (Heat, Lexington, Mass. pp. 1–12.

ROUSSEAU, R. (1988), Citation distribution of pure mathematics journals. In: L. EGGHE, R. ROUSSEAU (Eds), *Informetrics 87/88,* Elsevier Science Publishers B.V., pp. 249–260.

SCHUBERT, A., GLÄNZEL W. (1983), Statistical reliability of comparisons based on the citation impact of scientific publications, *Scientometrics*, 5 : 59–74.

SCHUBERT, A., GLÄNZEL, W. (1986), Mean response time. A new indicator of journal citation speed with application to physics journals, *Czechoslovakian Journal of Physics*, B 36 : 121–125.

SCHUBERT, A., GLÄNZEL, W., BRAUN, T. (1987), Subject field characteristic citation scores and scales for assessing research performance, *Scientometrics*, 12 : 267–291.

SCHUBERT, A., GLÄNZEL, W., BRAUN, T. (1989), Scientometric Datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major fields and subfields 1981-1985, *Scientometrics*, 16 : 3–478.

SEGLEN, P. O. (1997), Citations and Journal Impact Factors: Questionable indicators of research quality, *Allergy,* 52 : 1050–1056.

SMART, J. C., ELTON, C. F. (1982), Consumption factor scores of psychology journals, *Scientometrics*, 4 : 349–360.

TODOROV, R. (1983), Condensed matter physics journals, *Scientometrics*, 5 : 291–301.

TODOROV, R., GLÄNZEL, W. (1988), Journal citation measures: A concise review, *Journal of Information Science,* 14 : 47–56.

TOMER, C. (1986), A statistical assessment of two measures of citation: the impact factor and the immediacy index, *Information Processing and Management*, 22 : 251–258.

VAN LEEUWEN, TH. N., MOED, H. F., REEDIJK, J. (1997), *JACS* still topping *Angewandte Chemie*: Beware of erroneous impact factors, *Chemical Intelligencer*, July, 32–36.

WALLACE, D. P. (1986), The relationship between journal productivity and obsolescence, *Journal of the American Society for Information Science*, 37 : 136–145.

YANOVSKY, V. I. (1981), Citation analysis significance of scientific journals, *Scientometrics*, 3 : 223–233.

Received September 30, 2001.

Address for correspondence:
WOLFGANG GLÄNZEL
Bibliometrics Service
Library of the Hungarian Academy of Sciences
P. O. Box 1002, H–1245 Budapest, Hungary
E-mail: glanzw@helka.iif.hu