On indexing in the Web of Science
and predicting journal impact factor

Xiu-fang WU[1], Qiang FU[2], Ronald ROUSSEAU[3,4]

([1]*Journals of Zhejiang University SCIENCE (A&B), Zhejiang University Press, Hangzhou 310027, China*)

([2]*Zhejiang University Press, Hangzhou 310028, China*)

([3]*KHBO-Association K.U. Leuven, Industrial Sciences and Technology, Zeedijk 101-8400 Oostende, Belgium*)

([4]*K.U. Leuven, Steunpunt O&O Indicatoren, Dekenstraat 2, B-3000 Leuven, Belgium*)

E-mail: jzus@zju.edu.cn; fuqiang@zju.edu.cn; ronald.rousseau@khbo.be

Abstract: We discuss what document types account for the calculation of the journal impact factor (JIF) as published in the Journal Citation Reports (JCR). Based on a brief review of articles discussing how to predict JIFs and taking data differences between the Web of Science (WoS) and the JCR into account, we make our own predictions. Using data by cited-reference searching for Thomson Scientific's WoS, we predict 2007 impact factors (IFs) for several journals, such as *Nature*, *Science*, *Learned Publishing* and some Library and Information Sciences journals. Based on our colleagues' experiences we expect our predictions to be lower bounds for the official journal impact factors. We explain why it is useful to derive one's own journal impact factor.

Key words: WoS (Web of Science), JCR (Journal Citation Reports), Citation analysis, Predicted impact factors

doi:10.1631/jzus.B0840001   **Document code:** A

INTRODUCTION

The Science Citation Index (SCI) was launched in 1963 (covering 1961 data) and is widely used as a tool for the assessment of journals and even individual researchers. If it is to be used appropriately as part of a scientific research evaluation system by universities, funding organizations or governments, it is of the utmost importance to know and understand the specific indexing methods underlying this database. Yet, it is shown that this is not easy to find out. Eugene Garfield introduced the idea of an impact factor in 1955 (Garfield, 1955). This impact factor, however, referred to articles, not journals, and was not yet clearly defined. No mathematical formula for its calculation was proposed at that time. Several years later, he and Irving H. Sher created the journal impact factor (JIF). This index was designed for comparing journals regardless of their size and was a natural result of the establishment of the SCI (Garfield, 2001). The use of the IF as a visibility (quality?) measure is widespread because, according to (Garfield, 2006), it tends to correspond with scholars' view of the best journals in their own specialty.

The IF has often been criticized as unrepresentative or misleading (Rossner *et al.*, 2007). In recent years the number of articles, often editorials, discussing the IF is clearly on the rise (Bar-Ilan, 2008). Although there are many conflicting opinions about IFs, most people agree that the IF is only a numerical indicator of visibility and is such only weakly related to quality. Yet, 'quality' itself is a subjective notion. We note that the IF certainly does not reflect the quality of the peer review process to which a journal subjects submitted articles (Kurmis, 2003; Benítez-Bribiesca, 2002). Some editors seek to understand the IF calculation so that they can manipulate it to their journal's advantage (Jennings, 1998). Citation and

publication patterns differ between disciplines, so the IF is only meaningful when it is used to compare journals within the same discipline (Testa and McVeigh, 2004; Moed, 2005). An IF in one subject should never be compared with one in another; the relative ranking of journals within the same area is of greater importance, see e.g. (Rousseau and Smeyers, 2000) who provided an example where this has been done for local assessment of research groups. Great care is needed when using the IF as an evaluation tool (Zhang *et al.*, 2003). Particularly in comprehensive universities the evaluation of scientific research articles needs special care and should not be based on the IF of the journal in which the article is published (Seglen, 1994). Rousseau (2002) wrote that the quality of a journal is a multi-faceted notion: a simple IF can at best catch only one aspect. In this article we will focus on IFs, not on the Web of Science (WoS) and all its tools. Accuracy of citation counts in general (hence affecting IF calculations) is discussed in Chapter 13 of (Moed, 2005).

We recall that the Journal Citation Reports (JCR) published by Thomson Scientific evaluate and compare journals using citation data drawn from over 9100 scholarly and technical journals, and published meeting proceedings from more than 3300 publishers in over 60 countries (Journal Citation Reports 4.0, 2006). Although impressive, this number falls far short of the estimated total of more than 64000 academic/scholarly journals in existence, according to Morris (2007), while Brumback (2008) mentions 40000 journals of which about 15000 deserve to be called academic.

Also, we recall the definition of the JIF as used in the JCR. Journal *J*'s IF in the year *Y* is defined as: the ratio of the number of citations received in year *Y* by *all* documents published in journal *J* in the years *Y*–1 and *Y*–2 and the sum of the number of *citable* documents published in journal *J* in the years *Y*–1 and *Y*–2.

The term *citable* document will be discussed further in this paper.

The first author belongs to the editorial office of *Journals of Zhejiang University (A&B)*, two peer-reviewed journals (Zhang *et al.*, 2003) indexed by Thomson Scientific since 2007 and 2008 respectively, and publishing almost exclusively substantive research or review articles. It is in this function that we began the investigation reported here.

CONTENTS INDEXED BY THE WoS

A researcher's scientific attainments will generally be reflected and judged by his or her academic articles. As a journal provides a platform for communication with its own specific audience, it publishes more than just full scientific research papers, but also editorials, reviews, comments and short notes. Once a journal is covered by SCI, all items in it will be indexed by SCI databases, such as Science Citation Index-EXPANDED (SCI-E), Social Science Citation Index (SSCI), Index Chemicus (IC), Current Chemical Reactions-EXPANDED (CCR-E), irrespective of whether the articles appeared in regular issues, supplements or special issues. The indexers of the WoS distinguish the following publication types: *Article*, *Bibliography*, *Biographical-Item*, *Book Review*, *Correction-Addition*, *Database Review*, *Discussion*, *Editorial Material*, *Excerpt*, *Hardware Review*, *Item about an Individual*, *Letter*, *Meeting Abstract*, *Meeting Summary*, *News Item*, *Note*, *Reprint*, *Review*, *Software Review*. Some of these, however, are hardly used (*Meeting Summary* is a case in point) and will not be considered. More types are indexed in the Arts & Humanities Citation Index, but these too will not be considered.

In order to have an idea of the relative amount of publications of each type we searched Thomson Scientific's WoS and summarized all document types indexed On the WoS from 2004~2007, see Table A1 in Appendix.

Document types published and hence indexed the most are: *Article*, *Meeting Abstract*, *Editorial Material*, *Review*, *Letter* and *News Item*. In order to make a distinction between the WoS type *Article* and 'just' an article, the former will always be written in italic and with a capital A.

DOCUMENT TYPES AND CITABLE (SOURCE) ITEMS COUNTED IN THE IMPACT FACTOR CALCULATION

Not all indexed items, even those published in the most famous journals, are the outcome of scientific research. For example, parts of *Science* and *Nature* are of the (scientific) newspaper type. In these newspaper-type sections, news items, comments, and

short reports are published. Such items attract a large audience.

IF calculation is based on two elements: the numerator, which is the number of citations in the current year to *any items* published in a journal in the previous 2 years, and the denominator, which is the number of substantive articles (*citable* items) published in the same 2 years (Garfield, 1999). JCR's source data table shows that citable items in the JCR are further divided into *Articles* (i.e., research articles) and *Reviews*.

Generally speaking, an article is the direct report of an original research finding; a review consists of an overview, or sometimes a comment, on a certain problem by an expert in the field; letters (not letters to the editor) carry discussions of scientific problems. The data in these three document types is fairly complete; thus they are of more academic interest and usually receive more citations than other publication types. They are considered as citable (source) items when conducting the IF calculation. When a letters is primary a research report, it is considered as an *Article*, and becomes a source item.

In this article we analyzed all document types published in ten journals on Communication, Infor-mation Science and Library Science, including *Libri*, *Library Journal* (*Libr J*), *Scientist*, *Journal of Librarianship and Information Science* (*J Libr Inf Sci*), *Library & Information Science Research* (*Libr Inform Sci Res*), *Library Trends* (*Libr Trends*), *Scientometrics*, *Journal of Documentation* (*J Doc*), *Science and Engineering Ethics* (*Sci Eng Ethics*) and *Learned Publishing* (*Learn Publ*), trying to find out exactly which items are considered as source items.

For example, in *Learned Publishing*, the main sections are: guest editorial, editorial, articles, research articles, letter to the editor, industry developments, personal views, points of view, essay review, erratum, case study, book reviews, meeting report, etc. Research articles, case study and reports are considered to be of the *Article* type, reviews and essay review are *Review* type, and other sections are indexed as *Editorial Material*, etc. In *Library Journal*, the main sections are: Features, Bestsellers, Commentary, Departments, Info Tech, News, Reviews, and articles (very few) as source items are selected from Features, Commentary and Info Tech. Note that here Reviews are Book reviews, hence these reviews are not considered to be citable. *Libri* publishes almost exclusively articles and reviews. Table 1 showed the ten

Table 1 Ten journals' source data in the period of 2003~2007 [*]

Journal title	Document type	Citable items	Total items	Citable items percent	Citations of citable items	Citations of total items	Citation percent of citable items
Libri (quarterly)	Article (116), editorial material (2), reviews (1)	117	119	98.32%	60	60	100%
Libr J (semi-monthly)	Book review (26,975), letter (456), software review (73), biographical-item (21), editorial material (752), news item (221), bibliography (68), review (2), article (623), database review (80), correction (42)	625	29313	2.13%	129	239	53.97%
Scientist (monthly)	Editorial material (1,326), article (935), news item (728), letter (482), biographical-item (66), correction (29), book review (16), review (4), reprint (2)	939	3588	26.17%	320	725	44.14%
J Libr Inf Sci (quarterly)	Book review (99), article (75), editorial material (16), review (6)	81	196	41.33%	81	91	89.01%
Libr Inform Sci Res (quarterly)	Article (119), book review (80), editorial material (19), correction (2), software review (1), reprint (1), review (6)	125	228	54.82%	270	280	96.43%
Libr Trends (quarterly)	Article (238), editorial material (21), review (5), correction (1)	243	265	91.70%	210	220	95.45%
Scientometrics (monthly)	Article (544), letter (14), review (6), correction (2), editorial material (20), biographical-item (8), book review (4)	550	598	91.92%	1764	1796	98.22%
J Doc (double-monthly)	Book review (146), editorial material (23), review (8), correction (1), article (138), reprint (11)	149	327	44.65%	442	465	95.05%
Sci Eng Ethics (quarterly)	Article (233), letter (5), editorial material (51), book review (4), reprint (2), correction (2), review (5)	238	302	78.81%	309	322	95.96%
Learn Publ (quarterly)	Article (166), editorial material (40), biographical-item (4), correction (2), book review (53), letter (9), review (4)	170	278	61.15%	210	223	94.17%

[*]Data from Citation Report in Thomson Scientific's Web of Science (May 6, 2008)

journals' document types indexed by SCI in the period of 2003~2007.

Garfield (1994) pointed out that source items covered by SCI include not just original research papers, review articles and technical notes, but also letters, corrections and retractions, editorials, and other items. The latter items are important, have substantial impact, and provide useful links to scientific issues and controversies. Thomson Scientific also states that "The denominator contains a count of indexed *citable items*. Although all primary research articles and reviews (whether published in front-matter or anywhere else in the journal) are included, a citable item also includes substantive pieces published in the journal that are, bibliographically and bibliometrically, part of the scholarly contribution of the journal to the literature" (Pendlebury, 2007). Other items (*non-citable* items) including editorials, letters to the editor, news items, and meeting abstracts, etc. are not counted in the denominator of the IF as carried out for the JCR because they are not generally cited (Journal Citation Reports 3.0, 2004).

Many published comments on the IF calculation focus on the ratio of *citable* items and *non-citable* items and the contribution of non-citable items to the numerator of the IF (Jacsó, 2001; Editorial, 2005; Frandsen, 2008). Citations of so-called *non-citable* items contributing to a journal's IF are said to be 'for free' by Moed and van Leeuwen (1996). A recent study of this phenomenon is conducted by Golubic *et al.*(2008). They studied four journals in detail: the *New England Journal of Medicine* (*NEJM*) (in 2004, 81% of the published items were *non-citable*), *Nature* (63% *non-citable* items), *Anais da Academia Brasileira de Ciencias* (2% *non-citable* items) and *the Croatian Medical Journal* (31% *non-citable* items). They investigated the proportion of the original research results in these publications. For *Nature* they found original research data in only 94.7% of the *Articles*, while 9.5% of the editorial material and letters contained original research data (attracting quite a lot of citations). For *NEJM* 92.2% of the *Articles* contained original research data as did 7.2% of the editorial material and letters. Recalculating IFs using only published items containing original data and citations of these items they found that for all four journals the IF decreased by more than 10%, even by 32.3% for *NEJM*.

We also conducted the ten journals' citation analysis, and obtained the ratios of their citable items citations to total items citations in the period of 2003~2007. Table 1 shows that among these journals *Libr J* and *Scientist* have the lowest citable items percent namely 2.13% and 26.17%, respectively. Their ratios of citable items (articles and reviews) citations to total items citations are also lower than for the other journals. Other journals publish between 41% and 99% of citable items, receiving between 89% and 100% of their citations from these citable items. For *Learned Publishing* (Table 2), SCI indexed 47~70 items every year, with only 49%~74% *Articles* and *Reviews*. Its ratio of citable item citations to total items citations is about 90%. These data show that *Articles* and *Reviews* are the most cited items, accruing in almost all cases more than 90% of all citations.

An item is classified as a *Review* if it meets any of the following criteria: it cites more than one hundred references; it appears in a review publication or a review section of a journal; the word review or overview appears in its title; the abstract states that it is a review or survey (http://admin.isiknowledge.com/JCR/help/h_glossary.htm, 2008-02-23). So it is rather easy to distinguish reviews from all other indexed items.

The interpretation of the *Article* type, however, offers quite a challenge as it tends to differ from journal to journal.

Table 2 Journal source data of *Learned Publishing* (quarterly) in JCR years of 2003~2007[*]

JCR year	Citable items	Total items	Citable items percent	Citations of citable items	Citations of total items	Citation percent of citable items
2007	30	61	49.2%	9	11	81.8%
2006	29	47	61.7%	27	30	90.0%
2005	32	48	66.7%	33	36	91.7%
2004	38	52	73.1%	75	76	98.7%
2003	41	61	49.2%	53	11	94.6%

[*]Data from Citation Report in Thomson Scientific's Web of Science (Feb. 27, 2008)

WHY CALCULATE THE JIF?

Why are scientists, editors and administrators interested in JIFs, and, more importantly, why would they try to calculate them for themselves? We see four main reasons as follows:

1. It is for many purposes necessary to use another journal IF than the two-year synchronous one provided by Thomson Scientific. Data for calculating those other IFs can be found in the JCR. A general approach to all types of IFs is provided in (Frandsen and Rousseau, 2005; Ingwersen *et al.*, 2001).

2. Editors, responsible for journals not (yet) included in the JCR want to calculate their journal's IF and compare it to other journals who are included. This leads to concepts such as Stegman (1997; 1999)'s constructed IFs. Making predictions is also of interest for newly established journals which do not yet exist three years, the period needed to calculate the classical JIF. There are some exceptions, such as *Journal of Informetrics*, which started from Volume 1 in 2007 and soon was added to the WoS. "It will receive a 2008 IF which will be published in the 2009 JCR. This IF will be based on citations in 2008 to articles in the journal in 2007, the starting year" (Egghe, 2008). This means that the newly indexed journal beginning with Vol. 1 will have an official IF after two years, and this 'special' IF becomes: the ratio of the number of citations received in year Y by *all* documents published in journal J in the year $Y–1$ and the number of *citable* documents published in journal J in the year $Y–1$.

3. Editors want to know, or predict their journal' IF before its official release. This gives them more time to take counteraction measures, if necessary. If the results are favourable for their journal they may also prepare press releases in advance (Craig, 2007; Ketcham, 2007).

4. Verification of data and results is a normal part of scientific inquiry. For this reason scientometricians want to be able to check if the data and IFs provided in the JCR are indeed correct.

In view of this, we present this paper on the IF prediction. Using data from Thomson Scientific's WoS, we predicted IFs months before the release of the JCR. As a practical application we are able to evaluate the potential of specific journals—already in the collection or considered for inclusion—but for which in the previous year no IF exists.

CAN WE ESTIMATE IMPACT FACTOR BEFORE THE RELEASE OF JCR?

Thomson Scientific clearly stated that only original research and review articles are used in IF calculations (Journal Citation Reports on the Web v.4.0, 2008). Moreover, only when the article is published in its final form and indexed in WoS as a source item, it gets counted in the denominator for the IF calculation.

Trying to determine a journal's IF before the official release of the JCR needs the same procedure as for determining a journal's IF if this journal is not (yet) indexed by Thomson Scientific. Such an exercise has been done before by Spaventi *et al.*(1979), Sen *et al.*(1989) and in particular by Stegmann (1999) who published a number of so-called constructed IFs.

Craig (2007) points out that there are basically two procedures for estimating a journal's future IF. One is to search within the Web of Science and use the Citation Report feature. This approach, however, underestimates the true IF as it ignores errors leading to unmatched citations. The other one is to perform a cited-reference search. In this way, errors in volume, page numbers and so on are ignored. Typing errors in the name of the journal remain of course. As Thomson Scientific cleans the database before calculating IFs, less important typing errors are found by their approach (Pendlebury, 2007). Craig (2007) predicted that *The Journal of Sexual Medicine*'s 2006 IF would have a value somewhere between 3.8 and 4.6. It turned out that the official 2006 IF is 4.676, higher than the predicted IF. Based on the cited-reference search, Ketcham (2007) further proposed two methods to predict IFs through weekly collection of citation data from the WoS over the past 2 years. According to Table 4 in (Ketcham, 2007), we compared Ketcham's predicted 2006 IFs with the official ones released in July 2006, as shown in Table 3. Ketcham's method needs much work as many citation data must be collected and is time-consuming (it takes one half or even a whole year to track one or more journals' citation records weekly), but the error is relatively low. Hence the cited-reference search approach usually leads to a better prediction.

Table 3 Comparison of Ketcham's predicted 2006 IFs with the official IFs

Journal title	IF in 2006 JCR	Ketcham's predicted 2006 IF	Prediction error
Lab Invest	4.453	4.396	−1.28%
Modern Pathol	3.753	3.485	−7.14%
Am J Pathol	5.917	5.665	−4.26%
J Pathol	5.759	5.612	−2.55%
Am J Surg Pathol	4.144	4.165	0.51%
Hum Pathol	2.810	2.813	1.07%

When Rossner *et al.*(2007) examined the data for a number of journals published by the Rockefeller University Press, in the Thomson Scientific database the numbers did not add up. The total number of citations for each journal was substantially fewer than the number published in JCR website. This seems to be the case in most instances. Experts from Thomson Scientific said that the producers of the SCI and JCR have never claimed that these numbers should agree, since they result from two different analytic processes applied to the same underlying data. In keeping with the goal of producing an analysis appropriate to the specific task at hand—in one case search, navigation and study of individual articles and ad hoc groups of them, in the other case in-depth analysis of journals and journal ranking such different procedures are indeed required. See for instance the Thomson response to the Rossner editorial (Pendlebury, 2007) where ISI talks in detail about the discrepancies between IF data and WoS data, and explains why the JCR editorial team cannot simply rely on the self-identification of document types as presented in the journals. According to (Pendlebury, 2007) the reason seems to be that different derived (from the original raw data) databases are used for the WoS and for the determination of the JCR (and hence for the IF). But Thomson Scientific insists that there is essentially just one database. More data cleansing has been done for the JCR, leading to more accurate data, and in particular more citations received for journals.

We predicted IFs of the ten journals using the two methods hinted by Craig (2007), as shown in Table 4. The result shows that cites in 2007 to articles published in 2005 and 2006 are larger by Method 2 than those by Method 1, thus yielding IFs which are closer to the newly released official ones. So Method 2 seems a more feasible method for predicting IFs.

Table 4 Ten journals' IF predictions in JCR year

Journal title	2007 IF in JCR	Our predicted 2007 IF*	
		Method 1	Method 2
Libri	0.286	9/42=0.214 (−25.18%)	10/42=0.238 (−16.78%)
Libr J	0.295	38/234=0.162 (−13.30%)	64/234=0.274 (−7.12%)
Scientist	0.322	58/338=0.172 (−46.58%)	111/338=0.328 (1.86%)
J Libr Inf Sci	0.405	10/33=0.303 (−25.19%)	14/33=0.424 (4.69%)
Libr Inform Sci Res	0.870	39/54=0.722 (−17.01%)	45/54=0.833 (−4.25%)
Libr Trends	0.333	15/78=0.192 (−42.34%)	23/78=0.295 (−11.41%)
Scientometrics	1.472	322/248=1.298 (−11.82%)	387/248=1.560 (5.98%)
J Doc	1.309	89/68=1.309 (0%)	89/68=1.309 (0%)
Sci Eng Ethics	0.378	31/98=0.316 (−16.40%)	37/98=0.378 (0%)
Learn Publ	0.738	40/61=0.656 (−11.11%)	40/61=0.656 (−11.11%)

*Numbers of recent articles come from JCR; cites to recent articles in Method 1 from general search within the Web of Science and Citation Report, and in Method 2 from cited-reference search counted by hand; data in parentheses denote prediction errors

The JCR processing deadline usually is around mid-February following the JCR year, and the JCR is then published annually in June or July of that year. If using the predicted IFs, even taking the lag of Thomson Scientific's source data update into account, one can have an estimate for a journal's IF already at the end of February.

CASES OF *SCIENCE* AND *NATURE*

The journals *Science* and *Nature* are special cases. For this reason we discuss them in a separate section.

Nature publishes many types of sections such as: Editorials, Research Highlights, Journal Club, News, News Features, Business, Correction, Column, Commentaries, Correspondence, Books and Arts, Essay, News and Views, News and Views Q&A, Insight, Brief Communications, Brief Communications Arising, Commentary, Horizons, Feature, Articles, Letters, Technology Features, Corrigendum, Erratum, Retraction, Supplement, Focus, Naturejobs, Futures, Editor's Summary, Podcast, Authors, etc.

In *Science*, the main sections are: Special Issue, This Week in Science, Editorial, Editors' Choice,

News of the Week, News Focus, Letters, Books et al., Policy Forum, Perspectives, Review, Association Affairs, Brevia, Research Articles, Reports, and Technical Comments. According to *Science*, Research Articles are expected to present a major advance.

By analyzing the Table of Contents of *Nature* and *Science* and searching in the WoS, we found that for *Science*, Research Article, Association Affairs, Technical Comments, Reports, Brevia are considered to be of the *Article* document type, while for *Nature*, Article, Letters, Horizons, Feature, News Features, Technology Features, Brief Communications, Progress are *Article* document type. Note that in *Nature*, letters is of the *Article* type, but in *Science*, they are of the *Letter* type, as they discuss material published in *Science* in the last 3 months or issues of general interest. It seems that there is no exact i.e. globally applicable) definition of an article at Thomson Scientific. Worse, there are numerous incorrect *Article*-type designations in the WoS. Brumback (2008) for instance, mentioned that in 2005 two sets of meeting abstracts in the *Journal of Child Neurology* were classified as *Articles*. Moreover, we know of

cases where newly included journals are not included completely during the first year(s) leading to errors in the journal's IF (as some journal self-citations are missing).

Tables 5 and 6 show that from 2003 to 2007, SCI indexed more than 2500 items respectively from *Nature* and *Science* every year, with only 30% to 40% (between 800 and 1100 items) being *Articles* and *Reviews*. The ratio of citable items citations to total items citations is between 88% and 94%. We predicted IFs of *Nature* and *Science* using Method 1 suggested by Craig (2007), as shown in Tables 7 and 8. We found that the errors are more than 10%, agree well with that of Golubic *et al.*(2008).

DISCUSSION AND CONCLUSION

In recent years, more and more editors are focusing on their own journals' development, using citations in their analysis. It seems, however, to be a futile endeavour to try to exactly recreate the journal IFs (as provided by the JCR) using the WoS. In view of scientometricians' legitimate demand for data

Table 5 Journal source data of *Nature* (weekly) in JCR years of 2003~2007 from the WoS

JCR year	Citable items	Total items	Citable items percent	Citations of citable items	Citations of total items	Citation percent of citable items
2007	863	2679	32.2%	5614	6223	90.2%
2006	962	2733	35.2%	29001	31386	92.4%
2005	1041	2808	37.1%	57622	62678	91.9%
2004	937	2603	36.0%	70181	75984	92.3%
2003	1001	2590	38.6%	103980	110797	93.8%

Table 6 Journal source data of *Science* (weekly) in JCR years of 2003~2007 from the WoS

JCR year	Citable items	Total items	Citable items percent	Citations of citable items	Citations of total items	Citation percent of citable items
2007	890	2551	34.9%	5105	5707	89.5%
2006	886	2632	33.7%	22540	25346	88.9%
2005	933	2698	34.6%	47651	52319	91.1%
2004	924	2682	34.5%	78041	84704	92.1%
2003	926	2624	35.3%	97133	106644	91.1%

Table 7 *Nature* IF predictions in JCR year from the WoS

JCR year	IF in JCR	Predicted IF	Error
2007	28.751	25.948	−9.749%
2006	26.681	23.815	−10.74%
2005	29.273	25.143	−14.11%
2004	32.182	23.687	−26.40%
2003	30.979	23.714	−23.45%

Table 8 *Science* IF predictions in JCR year from the WoS

JCR year	IF in JCR	Predicted IF	Error
2007	26.372	23.104	−12.392%
2006	30.028	24.828	−17.32%
2005	30.927	27.020	−12.63%
2004	31.853	25.461	−20.07%
2003	29.781	26.824	−9.93%

verification this is not a good state of affairs. Moreover, errors are made by all actors involved (authors, publishers and Thomson Scientific's indexers). Editors and authors should make sure that reference lists are complete and accurate and that the correct abbreviations are used for the journals. Bad reference lists reduce IF values. Attempts of data manipulation by editors are another sad development, which does not help serious research evaluation exercises.

Using data from Thomson Scientific's WoS, we calculated IFs for a set of journals. If earlier experiences are confirmed our predicted IFs will be lower than the official ones. Yet we think that the predicted IFs can be used to estimate IFs at least four months before their official release in June or July. They will usually provide an underestimate, but, if this is known, no harm is done.

ACKNOWLEDGEMENT

References

Bar-Ilan, J., 2008. Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, **2**(1):1-52. [doi:10.1016/j.joi.2007.11.001]

Benítez-Bribiesca, L., 2002. The ups and downs of the impact factor: the case of *Archives of Medical Research. Archives of Medical Research*, **33**(2):91-94. [doi:10.1016/S0188-4409(01)00373-3]

Brumback, R.A., 2008. Worshipping false idols: the impact factor dilemma. *Journal of Child Neurology*, **23**(4):365-367. [doi:10.1177/0883073808315170]

Craig, I.D., 2007. The *Journal of Sexual Medicine*—impact factor predictions and analysis. *Journal of Sexual Medicine*, **4**(4i):855-858. [doi:10.1111/j.1743-6109.2007.00515.x]

Editorial, 2005. Not-so-deep impact: Research assessment rests too heavily on the inflated status of the impact factor. *Nature*, **435**(7045):1003-1004 [doi:10.1038/4351003a]

Egghe, L., 2008. Announcement: *Journal of Informetrics* is a Source Journal. [SIGMETRICS] Batch E-mail.

Frandsen, T.F., 2008. On the ratio of citable versus non-citable items in economics journals. *Scientometrics*, **74**(3):439-451. [doi:10.1007/s11192-007-1697-9]

Frandsen, T.F., Rousseau, R., 2005. Article impact calculated over arbitrary periods. *Journal of the American Society for Information Science and Technology*, **56**(1):58-62. [doi:10.1002/asi.20100]

Garfield, E., 1955. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, **122**(3159):108-111. [doi:10.1126/science.122.3159.108]

Garfield, E., 1994. The Concept of Citation Indexing: a Unique and Innovative Tool for Navigating the Research Literature. *Current Contents (print editions)*, January 3, 1994.

Garfield, E., 1999. Journal impact factor: a brief review. *CMAJ*, **161**(8):979-980.

Garfield, E., 2000. The use of *JCR* and *JPI* in measuring short and long term journal impact. *The Scientist*. Presented at Council of Scientific Editors Annual Meeting, May 9, 2000.

Garfield, E., 2001. Recollections of Irving H. Sher 1924-1996: polymath/information scientist extraordinaire. *Journal of the American Society for Information Science and Technology*, **52**(14):1197-1202. [doi:10.1002/asi.1187]

Garfield, E., 2006. The history and meaning of the journal impact factor. *Journal of the American Medical Association*, **295**(1):90-93. [doi:10.1001/jama.295.1.90]

Golubic, R., Rudes, M., Kovacic, N., Marusic, M., Marusic, A., 2008. Calculating impact factor: how bibliographical classification of journal items affects the impact factor of large and small journals. *Science and Engineering Ethics*, **14**(1):41-49. [doi:10.1007/s11948-007-9044-3]

Ingwersen, P., Larsen, B., Rousseau, R., Russell, J.M., 2001. The publication-citation matrix and its derived quantities. *Chinese Science Bulletin*, **46**(6):524-528.

Jacsó, P., 2001. A deficiency in the algorithm for calculating the impact factor of scholarly journals—The journal impact factor. *Cortex*, **37**(4):590-594. [doi:10.1016/S0010-9452(08)70602-6]

Jennings, C., 1998. Citation data: the wrong impact? *Nature Neuroscience*, **1**(8):641-642. [doi:10.1038/3639]

Journal Citation Reports 3.0, 2004. Http://scientific.thomson-reuters.com/support/faq/wok3new/jcr3/, uploaded on 2004-08-09, accessed on 2008-01-10.

Journal Citation Reports 4.0, 2006. Http://scientific.thomson-reuters.com/support/faq/wok3new/JCR4/, uploaded on 2006-02-21, accessed on 2008-02-23.

Journal Citation Reports on the Web v.4.0, 2008. Http://scientific.thomsonreuters.com/media/scpdf/jcr4_sem_0305.pdf, accessed on 2008-05-28.

Ketcham, C.M., 2007. Predicting impact factor one year in advance. *Laboratory Investigation*, **87**:520-526. [doi:10.1038/labinvest.3700554]

Kurmis, A.P., 2003. Understanding the limitations of the journal impact factor. *The Journal of Bone and Joint Surgery (American)*, **85**:2449-2454.

Moed, H.F., 2005. Citation Analysis in Research Evaluation. Springer, Dordrecht.

Moed, H.F., van Leeuwen, T., 1996. Impact factors can mislead. *Nature*, **381**(6579):186. [doi:10.1038/381186a0]

Morris, S., 2007. Mapping the journal publishing landscape. *Learned Publishing*, **20**(4):299-310. [doi:10.1087/09531]

5107X239654]

Pendlebury, D.A., 2007. Thomson Scientific Corrects Inaccuracies in Editorial. Article titled "Show me the data", Journal of Cell Biology, Vol. 179, No.6, 1091-1092, 17 December 2007 (doi:10.1083/jcb.200711140) is Misleading and Inaccurate. Http://scientific.thomson.com/citationimpactforum/8427045/, accessed on 2008-03-25.

Rossner, M., Hill, E., van Epps, H., 2007. Show me the data. *Journal of Experimental Medicine*, **204**(13):3052-3053. [doi:10.1084/jem.20072544]

Rousseau, R., 2002. Journal evaluation: technical and practical issues. *Library Trends*, **50**(3):418-439.

Rousseau, R., Smeyers, M., 2000. Output financing at LUC. *Scientometrics*, **47**(2):379-387. [doi:10.1023/A:1005651429368]

Sen, B.K., Karanji, A., Munshi, U.M., 1989. A method for determining the impact factor of a non-SCI journal. *Journal of Documentation*, **45**(1):139-141. [doi:10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y]

Seglen, P.O., 1994. Causal relationship between article citedness and journal impact. *Journal of the American Society for Information Science*, **45**:1-11.

Spaventi, J., Tudor-Silovic, N., Maricic, S., Labus, M., 1979. Bibliometrijska analiza znanstvenih casopisa iz Jugoslavije (Bibliometric analysis of scientific journals from Yugoslavia). *Informatologia Yugoslavica*, **11**:11-23.

Stegmann, J., 1997. How to evaluate journal impact factors. *Nature*, **390**(6660):550. [doi:10.1038/37463]

Stegmann, J., 1999. Building a list of journals with constructed impact factors. *Journal of Documentation*, **55**(3):310-324. [doi:10.1108/EUM0000000007148]

Testa, J., McVeigh, M.E., 2004. The Impact of Open Access Journals: a Citation Study from Thomson ISI. Thomson-ISI.

Zhang, Y.H., Yuan, Y.C., Jiang, Y.F., 2003. An international peer-review system for a Chinese scientific journal. *Learned Publishing*, **16**(2):91-94. [doi:10.1087/095315103321505557]

APPENDIX

Table A1 Document types indexed on the WoS in the years 2004~2007

Document type	Items indexed by SCI-E				Items indexed by SSCI			
	2004	2005	2006	2007	2004	2005	2006	2007
Article	>100000	>100000	>100000	>100000	76781	86118	94055	93912
Bibliography	106	92	73	70	63	75	68	52
Biographical-Item	4007	4011	4345	4112	744	780	880	799
Book Review	3526	3443	3499	3223	28936	28548	28599	25706
Correction, Addition	7998	9166	9834	9347	625	703	707	710
Database Review	5	7	17	17	21	12	16	10
Editorial material	52135	55834	59937	57757	11375	12702	13890	12987
Hardware review	17	8	7	2	17	8	7	2
Item about an individual	4007	4011	4345	4112	4007	4011	4345	4161
Letter	34981	35651	36604	35881	34981	35651	36604	36150
Meeting Abstract	>100000	>100000	>100000	>100000	>100000	>100000	>100000	>100000
Meeting summary		5				3		
News Item	21183	22916	23376	21379	21183	22916	23376	21630
Reprint	559	590	621	443	76	130	96	77
Review	39091	41934	46796	45159	4758	5573	6256	5908
Software Review	158	132	158	133	57	43	28	52