

A DEFICIENCY IN THE ALGORITHM FOR CALCULATING THE IMPACT FACTOR OF SCHOLARLY JOURNALS: THE JOURNAL IMPACT FACTOR

Péter Jacsó

(Department of Information and Computer Science, University of Hawai'i at Mānoa,
Honolulu)

The Journal Citation Reports ® (JCR), published by the Institute for Scientific Information, annually ranks more than 6,500 science and social science journals, by calculating their impact factors (and many other useful measures) from the citations they receive from journals monitored by ISI in the Science Citation Index (SCI), Social Science Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI) databases that process about 25 million citations annually.

As Eugene Garfield, the founder of ISI, the inventor of the journal impact factor, and the deservedly most cited information scientist defines it: “[t]he annual *JCR* impact factor is a ratio between citations and recent *citable* items published. Thus, the impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years, by calculating the ratio between the number of citations received and the number of *citable* articles published” (Garfield, 1994 [emphases added]). For example, the impact factor of *Cortex* was determined by the algorithm shown in Figure 1 in the 1999 edition of *JCR* (the most current as of this writing).

THE DEFICIENCY IN THE ALGORITHM

The journal impact factor is a good measure insofar as it is tangible and reproducible, but it is not perfect and must not be taken at face value. There are various reasons for the imperfections of the impact factor; I discuss here only the one that is built into the algorithm.

ISI determines the denominator of the algorithm by counting only items that it considers “citable”. These include articles, notes, and reviews of the literature. Reviews of books, software, hardware and databases, editorials, corrections, letters to the editor, and other editorial materials, are not counted among the citable items. This restriction is not apparent when details are displayed of how the impact factor was calculated for a particular journal.

On the contrary, the algorithm that is displayed uses the same term “article” for both the nominator and the denominator, even though the latter includes only the items that are assigned the document types: review, note or article. The ISI Glossary further blurs the distinction in its definition of the impact factor that it

Impact Factor Calculation			
Journal:	CORTEX		
Impact Factor:	1.310		
Cites in 1999 to articles published in:	1997 =		104
	1998 =		44
	97 + 98 =		148
Number of articles published in:	1997 =		54
	1998 =		59
	97 + 98 =		113
Calculation:			
	$\frac{\text{Cites to recent articles}}{\text{Number of recent articles}}$	=	$\frac{148}{113} = 1.310$

Fig. 1 – The impact factor of Cortex in the 1999 edition of JCR.

defines as “[t]he number of current citations to articles published in a specific journal in a two year period divided by the *total number of articles* published in the same journal in the corresponding two year period” [emphasis added]. There are no definitions for the document types assigned by ISI, even though it is a crucial component in the algorithm. The denominator is almost always smaller, and quite often much smaller than the total number of articles published in the journal.

The difference between the total number of articles (at least those for which ISI creates a record) and the “citable” articles as calculated by ISI is significant. For example, 22% of the items in the Science Citation Index belong to the “non-citable” category. In the Social Science Citation index these non-citable items account for 46% of the total items, and there are even more (70.5%) in the Arts & Humanities Citation Index (which has no JCR counterpart).

“NON-CITABLE” ITEMS ARE CITED

Importantly, documents that the impact factor algorithm excludes from the denominator count are, in fact, cited, and the documents citing them are counted for the nominator. Moed and van Leeuwen illustrated the distortion that this policy may cause. None of the journals that they analyzed caused such extreme distortion as *Contemporary Psychology*, which had the highest impact factor ranking in the 1998 JCR (Jacsó, 2000). As *Contemporary Psychology* has only

book reviews, corrections and letters to the editor it should have not been included in the JCR stable of journals as its denominator must be zero in the spirit and practice of the impact factor algorithm.

As it turned out, three book reviews were mistakenly classified by ISI as articles. The journal received 30 citations for its book reviews, therefore its impact factor was 10, propelling the journal into the #1 position. Hopefully, ISI will remove this journal from the JCR stable, but this does not solve the problem in general.

This deficiency in the algorithm does not have a direct effect on journals such as *Cortex*, which publish only very few editorials, book reviews, corrections, and letters to the editor each year. However, it indirectly has a negative effect on the ranking of these journals because the algorithm favors journals having a larger proportion of non-citable items that do get cited.

As an example from the Neuroscience category, consider the impact factors of two journals, *Cortex* and *Pain Forum* (of the American Pain Society) that were constructed for this article to estimate their impact factor for JCR 2000 (to be published later in 2001). It is to be emphasized that the data for this article were derived by searching and analyzing results from the citation databases through a fairly complex procedure, i.e. the analysis does not use the official raw data of JCR 2000, which are not yet available.

In my estimate of JCR 2000 (shown in Figure 2), *Cortex*'s impact factor is predicted to increase slightly from 1.310 in 1999 to 1.427, while *Pain Forum* will see a dramatic increase from 2.700 in 1999 to 4.320. The difference between these two journals is caused precisely by the flaw in the impact factor algorithm, which handicaps more than ever *Cortex* and other journals that predominantly publish items that are considered "citable" by ISI. This is not muckraking, using yet another extreme example. There are many other journals in the Neurosciences category and in the entire JCR stable that benefit in their ranking from the deficiency of the algorithm.

For *Pain Forum* ISI classified 25 of the total number of publications as "citable" for 1998-1999. In reality, of the 25 citable items, 16 received a total of 58 citations, and the "non-citable" items received a total of 47 citations, yielding

Components of the impact factor algorithm	<i>Cortex</i>	<i>Pain Forum</i>
Cites in 1999 to articles published in 1997+1998	148	54
Number of "citable" items published in 1997+1998	113	20
Impact factor in JCR 1999	1.310	2.700
Actual number of items included in ISI databases for 1997+1998	115	109
Impact factor for JCR 1999 (as suggested by Jacsó)	1.287	0.495
Expected cites in 2000 to articles published in 1998+1999	157	105
Expected number of "citable" items published in 1998+1999	110	25
Expected impact factor in JCR 2000	1.427	4.200
Actual number of items included in ISI databases for 1998+1999	113	101
Impact factor for JCR 2000 (as suggested by Jacsó)	1.389	1.040

an unrealistically high impact factor for *Pain Forum*. The subset of its 1999 publications (not detailed in the aggregated table) sheds light on the source of distortion particularly well. The “citable” publications received 20 citations, while “non-citable” publications received 33 citations. Many other journals that have a mixture of “citable” and “non-citable” items prove not only that their “non-citable” subset gets significant citations, but that this “non-citable” subset may garner more citations than the “citable” subset.

AN AGGRAVATING FACTOR IN THE DISTORTION

Moed and Van Leeuwen suggest that there be an impact factor for each document type. This is reasonable, but it is a daunting task and is hindered by the fact that ISI is not consistent in classifying documents by type. For the 1999 issues of *Pain Forum* ISI assigned 5 items to the document type “article”. In fact, these items were commentaries and should have been classified as “editorial material” as in previous years. Had ISI been consistent, there would have been only 10 citable items for 1999 (just as in 1998, and 1997), and the impact factor would have increased to 5.40 (108/20), further distorting its impact factor. Such inconsistent classification by document type is prevalent in ISI databases, therefore the impact factor by document type would be distorted, too. When ISI assigns to an article or literature review a document type that makes it “un-citable” for the algorithm, it boosts the journal’s impact factor, and more importantly its ranking among its peer journals.

SUGGESTION

I suggest that we take the bull by the horns and use in the denominator the total number of published items for which ISI creates a record. This would minimize the distortion caused by applying the filter of “citable” items, which is further aggravated by inconsistencies in classification. As a compromise, this modified impact factor, which is very easy to calculate, should be published along with the traditionally calculated impact factor to let JCR users make their educated choice, and that the journals be accordingly re-ranked. In our example, this would very slightly reduce the impact factor of *Cortex*, but would drastically reduce the impact factor of *Pain Forum*, which benefited unjustly from the flaw of the algorithm and the inconsistency of the document type assignment. In the mid-range of the neurosciences journals, the suggested algorithm would considerably reshuffle the rankings.

In an interesting twist, it can be easily predicted that the impact factor and rank of *Pain Forum* will further increase in the 2001 JCR (to be published in 2002). This journal was replaced by the *Journal of Pain* in 2000. This means that the denominator (the number of “citable” items) will not increase, but the numerator certainly will by virtue of the fact that the 2001 issues of journals monitored by ISI will give additional citations to items published in the 1999 volume of *Pain Forum*, thus raising its impact factor.

REFERENCES

- GARFIELD, E. The impact factor. *Current Contents®* (25): 3-7, 20 June 1994 (also available at <http://www.isinet.com/isi/hot/essays/journalcitationreports/7.html>)
- ISI Glossary. <http://sunweb.isinet.com/isi/search/glossary/>
- JACSO, P. Savvy searching: The number game. *Online Information Review* 24: 180-183, 2000.
- JACSO, P. The impact of inconsistent document type assignment on the journal impact factors (in preparation, 2001).
- MOED, H.F., and VAN LEEUWEN, T.H. Improving the accuracy of [the] Institute for Scientific Information's journal impact factors. *Journal of the American Society for Information Science* 46: 461-467, 1995.