# What does the Web of Science five-year synchronous impact factor have to offer?①

Ronald Rousseau[1,2,3]*

[1] Industrial Sciences and Technology, KHBO (Association K.U. Leuven), Zeedijk 101, B-8400 Oostende, Belgium
[2] IBW, University of Antwerp, Venusstraat 35, B-2000 Antwerpen, Belgium
[3] Department of Mathematics, K.U. Leuven, Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium

Received Sept. 27, 2009
Accepted Nov. 7, 2009

**Abstract** With a random sample of 10 JCR (Science) subject areas it is shown that the 2-year and the 5-year impact factor of journals lead statistically to the same ranking per category. Yet in a majority of cases, the 5-year impact factor is larger than the 2-year one.

**Keywords** Synchronous impact factor, Wilcoxon matched-pairs, Signed-ranks test

## 1 Introduction

Since Thomson Reuters' Web of Knowledge included a five-year synchronous impact factor in its 2007 edition of the *Journal Citation Reports* (JCR), this addition has been generally treated favourably[1]. The definition of this impact factor is modelled after that of the standard Garfield-Sher two-year impact factor. For the year *Y*, and a given journal *J*, it is defined as:

$$IF_5(Y) = \frac{\sum_{k=1}^{5} CIT(Y-k, Y)}{\sum_{k=1}^{5} PUB(Y-k)}$$

where $CIT(Y-k,Y)$ denotes the number of citations received by journal *J* in the year *Y*, referring to publications in *J* during the years $Y-5$ to $Y-1$. The corresponding number of publications (in the year $Y-k$) is denoted as $PUB(Y-k)$.

To the best of our knowledge, the first formal definition of this, a similar synchronous impact factor, was provided by Rousseau in 1987 during the 1st

① This work is supported by the National Natural Science Foundation of China (Grant No. 70673019).

* Correspondence could be addressed to Ronald Rousseau (E-mail: ronald.rousseau@khbo.be). The author would like to express his sincere thanks to Mia Hubert (K.U. Leuven) for her statistical advice.

CJLIS
Vol. 2 No. 3, November 2009
pp 1–7
National Science Library, Chinese Academy of Sciences

International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval (nowadays called the ISSI conferences)[2]. In that article Rousseau studied different synchronous impact factors for mathematics journals. Yet, we have no doubt that Garfield and Sher experimented with other than the 2-year impact factor, probably including the 5-year one. Certainly, Garfield published comparisons between long-term (7-year and 15-year) and the short-term (2-year) impact factors[3,4]. On the one hand, Garfield observed that most top journals stay top journals, no matter how their impact is measured. Cases in point are *Cell*, *New England Journal of Medicine*, *Proceedings of the National Academy of Sciences USA*, *Nature*, *Science* and the *Journal of Experimental Medicine*. On the other hand, he noted that some journals show dramatic changes in impact rankings (almost all letters journals). By studing chemical journals, Vinkler[5] also compared the 5-year synchronous impact factor with the two-year and the ten-year synchronous impact factors (he refers to these impact factors as asynchronized impact factors), and found an excellent agreement between impact factors calculated for these different time windows based on a comparison of Pearson correlation coefficients.

## 2 Problem

Recently, Leydesdorff[6] studied if new indicators, such as the *h*-index, PageRank and the Scimago Journal Ranking really contribute to our knowledge of journal influence. More precisely, he was interested to find out if the new and the existing indicators measure new dimensions of the citation network. It is remarkable that in Leydesdorff's Fig. 2, a scatterplot of the two main components (productivity and impact), the 2-year impact factor and the 5-year impact are situated almost on top of one another, heavily loading on the impact dimension. We wondered if these two impact factors were completely equivalent. If this were the case, the 5-year one would be superfluous.

## 3 The basic model

In 2001, Rousseau et al.[7] described the basic synchronous citation model and proved that in this model the three-year synchronous impact factor is expected to be larger than the two-year synchronous impact factor. Under the term 'basic model' we mean a synchronous citation curve that is unimodal and has a mode at year two or three after publication. The basic model further assumes that the number of publications per year is not decreasing in time. By using the Chinese Science Citation Database (CSCD) as their research database, Rousseau et al. observed several discrepancies with this basic model and conjectured similar discrepancies in other databases. They also pointed out that the determination of a citation window (2, 3 or 5 years) is important in research evaluation assessments.

National Science Library, Chinese Academy of Sciences

What does the basic model have to say about the relation between the two- and the five-year synchronous impact factor? Comparing the two- and the five-year impact factor we find:

$$IF_2(Y) \leq IF_5(Y)$$
$$\Leftrightarrow$$
$$\frac{\sum_{k=1}^{2} CIT(Y-k,Y)}{\sum_{k=1}^{2} PUB(Y-k)} \leq \frac{\sum_{k=1}^{5} CIT(Y-k,Y)}{\sum_{k=1}^{5} PUB(Y-k)}$$
$$\Leftrightarrow$$
$$\left(\sum_{k=1}^{2} CIT(Y-k,Y)\right) \cdot \left(\sum_{k=3}^{5} PUB(Y-k)\right) \leq \left(\sum_{k=3}^{5} CIT(Y-k,Y)\right) \cdot \left(\sum_{k=1}^{2} PUB(Y-k)\right)$$

Assuming that the number of publications is constant, we find:

$$IF_2(Y) \leq IF_5(Y) \Leftrightarrow \frac{3}{2} \leq \frac{\sum_{k=3}^{5} CIT(Y-k,Y)}{\sum_{k=1}^{2} CIT(Y-k,Y)} \qquad (1)$$

The assumptions of the basic model say nothing about how fast the synchronous citation curve decreases. Intuitively, it seems that for a fast rising and fast decreasing field Inequality (1) is unlikely to hold, especially if the number of publications is not constant but increasing (making the left hand side larger). For slow moving disciplines it becomes more probable that Inequality (1) is satisfied.

## 4 Method

We consider the following two questions:

- Is, in a given subject area, the ranking of the journals based on the standard, 2-year impact factor, the same as the ranking based on the 5-year one?
- Is the average or median 2-year impact factor for a given subject area the same as the average or median 5-year impact factor?

We randomly chose 10 JCR (Science Edition) subject areas (using a random number generator), with each containing 40 journals at least. Journals with incomplete data were removed from the resulting lists.

The first question can be answered by calculating Spearman rank correlations. The Spearman rank correlation coefficient determines the degree to which a monotonic relationship exists between two variables (cf. Sheskin's Test 29)[8]. The second hypothesis regarding the means can be tested by calculating, for each journal, the difference between the 5-year and the 2-year impact factor, and by testing if this difference has average zero. According to Sheskin's Test 17[8], the *t*-test for two dependent samples is only valid if each of the underlying populations is normal, and with the same variance. Impact factors are (almost) never normally distributed and the same holds for the difference set (which we also checked on their normality). Hence, in an attempt to obtain data closer to a normal distribution we took logarithms. Yet, although the resulting distributions looked much more like a normal distribution, in most cases, also here, the hypothesis of normally distributed sets was rejected. We conclude that this test is not applicable. When the assumptions of the *t*-test for two dependent samples are violated, Wilcoxon's matched-pairs signed-ranks test offers a powerful solution (cf. Sheskin's Test 18)[8]. Its null-hypothesis is that the median of difference scores is zero. Details of this test are provided in Appendix I. If the null hypothesis is rejected, one can conclude that the two dependent samples represent different populations.

We finally note that the Mann-Whitney *U*-test can also not be applied[9]. The Mann-Whitney *U*-test as described in Sheskin's Test 12[8] is based on the assumption that the two samples are independent. This assumption is clearly violated here.

## 5 Results

The 10 JCR subject categories that were randomly chosen are shown in Table 1. Data related to the journals in these categories were collected from the JCR 2007 file (Science Edition).

Table 1 List of JCR subject categories used in this study

| Number | JCR subject categories |
|---|---|
| 1 | Biotechnology & Applied Microbiology |
| 2 | Cardiac & Cardiovascular Systems |
| 3 | Computer Science, Information Systems |
| 4 | Engineering, Multidisciplinary |
| 5 | Immunology |
| 6 | Instruments & Instrumentation |
| 7 | Metallurgy & Metallurgical Engineering |
| 8 | Ophthalmology |
| 9 | Physics, Multidisciplinary |
| 10 | Radiology, Nuclear Medicine & Medical Imaging |

In each subject category rankings between the 2-year and 5-year impact factor correspond well to a large extent, as witnessed by the Spearman rank correlation coefficients shown in the second column of Table 2. No statistical test is necessary as results speak for themselves. For the Wilcoxon test, using the normal approximation (Appendix I), the null hypothesis of zero median difference value is rejected in 7 cases of the 10 (test on 5% level; see Table 2, last column). Consequently, we

National Science Library, Chinese Academy of Sciences

considered a one-sided test to see if the median value is larger than zero, indicating that the 5-year impact factor is larger than the 2-year impact factor. This holds in 6 cases of the total 10 subject categories.

Table 2 Subject category rankings test of the 10 chosen JCR subject categories

| Subject area | Spearman | $\|z$-value$\|$ | $p$-value of $z$-test |
|---|---|---|---|
| Biotechnology & Applied Microbiology | 0.97 | 4.383 (+) | < 0.0001 |
| Cardiac & Cardiovascular Systems | 0.97 | 1.173 (−) | 0.241 |
| Computer Science, Information Systems | 0.96 | 6.627 (+) | < 0.0001 |
| Engineering, Multidisciplinary | 0.92 | 5.355 (+) | < 0.0001 |
| Immunology | 0.97 | 0.689 (−) | 0.491 |
| Instruments & Instrumentation | 0.98 | 3.615 (+) | 0.0003 |
| Metallurgy & Metallurgical Engineering | 0.96 | 4.847 (+) | < 0.0001 |
| Ophthalmology | 0.97 | 2.896 (+) | 0.0034 |
| Physics, Multidisciplinary | 0.96 | 1.629 (−) | 0.103 |
| Radiology, Nuclear Medicine & Medical Imaging | 0.95 | 3.866 (+) | 0.0001 |

Note: A "+" sign in the third column indicates that $R^+ > |R^-|$ (see Appendix I for an explanation of the notation).

Considering the *t*-tests, only for Instruments & Instrumentation can the hypothesis that the original sets are normally distributed not be rejected (10% level, using the Kolmogorov-Smirnov and Anderson-Darling tests as provided by the statistical software package Statgraphics Plus). For this case, the null-hypothesis of equal means is rejected at the 5% level. Taking logarithms of the impact factors brings the distribution closer to a normal one. Yet, even then only for such subject categories as Metallurgy & Metallurgical Engineering, Cardiac & Cardiovascular Systems, and Computer Science, Information Systems the hypothesis of a normal distribution (that the original data sets are log-normally distributed) cannot be rejected. However, even after this transformation the hypothesis of equal means is rejected in the three cases. Finally, we mention that the average over all included journals (737 in total) of the ratio (5-year impact factor)/(2-year impact factor) is equal to 1.144 (standard deviation: 0.352). That means that the 5-year impact factor is 14% larger than the standard two-year impact factor on average.

## 6 Conclusion

Journal indicators are most useful for comparing journals in the same subject category. In this respect, the 2-year and the 5-year impact factor lead statistically to the same ranking. Yet, it seems that in many cases, but not always, the 5-year impact factor is larger than the 2-year one. It would be interesting to characterize those JCR subject areas for which the 5-year impact factor is generally (median value) larger than the 2-year one, and those for which this is not the case.

Comparing the average impact factors using the standard *z*-test or *t*-test is for several reasons not a good idea. In particular, impact factors are in general not normally distributed, making a standard *t*-test invalid. Although the lognormal distribution fits better than the normal distribution, even a fit of the lognormal distribution is often rejected.

In a follow-up study we will compare 2-year and 5-year impact factors for ranges within one large, subject area [10].

## References

1 Della Sala, S. & Grafman, J. Five-year impact factor. Cortex, 2009, 45:911.

2 Rousseau, R. Citation distribution of pure mathematics journals. In Egghe, L. & Rousseau, R. Eds. Informetrics 87/88 . Amsterdam: Elsevier, 1988, 249–262.

3 Garfield, E. Long-term vs. short-term journal impact: Does it matter? The Scientist, 1998, 12(3):11–12.

4 Garfield, E. Long-term vs. short-term impact: Part II. Cumulative impact factors. The Scientist, 1998, 12(14):12–13.

5 Vinkler, P. Ratio of short term and long term impact factors and similarities of chemistry journals represented by references. Scientometrics, 1999, 46:621–633.

6 Leydesdorff, L. How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology, 2009, 60(7):1327–1336.

7 Rousseau, R., Jin, B. H. & Yang, N. H. et al. Observations concerning the two- and three-year synchronous impact factor, based on the Chinese Science Citation Database. Journal of Documentation, 2001, 57(3):349–357.

8 Sheskin, D. J. Handbook of parametric and nonparametric statistical procedures. Boca Raton (FL): Chapman & Hall, 2007.

9 Egghe, L. & Rousseau, R. Introduction to Informetrics. Amsterdam: Elsevier, 1990.

10 The STIMULATE 9 Group. The 5-year synchronous impact factor for large JCR subject areas. Preprint (2009).

## Appendix I: The Wilcoxon matched-pairs signed-ranks test

As the Wilcoxon matched-pairs signed-ranks test is not often used in our field[a)–c)], we provide, for the readers' convenience, a full description taken from Sheskin's Test 18.

1) The differences between the scores of the two dependent samples (here the value of the 2-year and of the 5-year impact factor) are calculated.
2) One considers the absolute values of these differences. Zero values are removed. The number of remaining values is denoted as *n*.

Notes: a) Antelman, K. Do open-access articles have a greater research impact? College & Research Libraries, 2004, 65(5):372–382.

b) Jansen, B. J. & McNeese, M. D. Evaluating the effectiveness of and patterns of interactions with automated searching assistance. Journal of the American Society for Information Science and Technology, 2005, 56(14):1480–1503.

c) Rodriguez, V., Janssens, F. & Debackere, K. et al. On material transfer agreements and visibility of researchers in biotechnology. In Torres-Salinas, D. & Moed, H. F. Eds. Proceedings of ISSI 2007. Madrid: CSIC, 2007, 659–671.

3) These absolute values are ranked (from smallest to largest).
4) The original sign of the difference is added to the rank.
5) Sums are taken for all positive ranks ($R^+$) and for all negative ranks ($R^-$) separately.
6) The absolute value of the smaller of these two sums is called the Wilcoxon $T$ test statistic, denoted as $T$.
7) This $T$-value is compared with the values provided in tables of critical values for Wilcoxon's matched-pairs signed-ranks test. For large $n$ there exists a normal approximation. The $z$-value is:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

This value is always negative, but this plays no role when performing a non-directional (two-sided) test. When a directional alternative hypothesis is employed, the null hypothesis can only be rejected if the directional alternative hypothesis that is consistent with the data is supported.