Arch. Immunol. Ther. Exp., 2009, **57**, 39–43
PL ISSN 0004-069X
DOI 10.1007/s00005-009-0006-0

Rank-normalized journal impact factor as a predictive tool*

Grzegorz Racki

Institute of Paleobiology, Polish Academy of Sciences, Warszawa, Poland

Received: 2008.11.25, **Accepted:** 2008.12.12, **Published online:** 2009.02.14

Abstract

Citation data accumulated on articles from the top and bottom 25% of impact factor (IF)-ranked international journals are compared using 59 international geoscience journals from 1998 and 378 Polish geological papers from 1989–1994. There is a minor risk of being uncited when results are published in high-IF periodicals as the average non-citation rate is 0.88% over a10-year period in this not very rapidly developing scientific discipline. Similarly, the established error levels in the prognosis of expected citation success versus failure based on the extreme IF quartiles as an evaluation tool is low (at most 12.5%). Thus the application of the rank-normalized journal IF as a proxy of real citation frequency and, accordingly, as a predictive tool in the *a priori* qualification of recently published publications is a rational time- and cost-saving alternative (or at least a significant supplement) to traditional informed peer review. Blanket criticism of using IF for decisions in research funding is therefore at least partly exaggerated.

Key words: impact factor, citation, scientific quality, research funding.

Abbreviations: ISI – Institute for Scientific Information in Philadelphia, IF – journal impact factor, r – correlation coefficient, n – number of articles, m – median citation value.

Corresponding author: Grzegorz Racki, Institute of Paleobiology, Polish Academy of Sciences, Twarda 51/55, 00-818 Warszawa, Poland, e-mail: racki@twarda.pan.pl

INTRODUCTION

The evaluation of scientific literature based on the degree to which publications are cited by others has become one of the more effective quantitative ways for measuring research performance. However, many pitfalls, biases, and constraints are encountered in the world-wide multidisciplinary citation databases (such as the Science Citation Index Extended) established by the Institute for Scientific Information (ISI) in Philadelphia, presently owned by Thomson Reuters. A key controversy is the rise in the ISI's journal impact factor (IF), a measure of the frequency with which an "average paper" from a journal has been cited in a particular period. It is frequently expected that the readily available IF sufficiently evaluates publication quality and is increasingly used to judge the value of scientific output (as publication "citedness"). The surrogate metrics of citation frequency also influences funding decisions in several countries, including Poland since 1998. Such a simplified bibliometric approach to academic evaluation for tenure or in assessing grant applications ("the poor man's citation analysis", van Raan in Adam 2002) has created a wide-ranging debate and is often rightly criticized (see different views in Fava and Ottolini 2000; Garfield 1983; Garfield 1996; Garfield 2000; Garfield 2003; Garfield 2006; Hoeffel 1998; Holden et al. 2006; Kabala 1998; Saha et al. 2003; Seglen 1994; Seglen 1997; Wróblewski 2008; Żylicz 1999).

Generally, about 15% of the citable items in a journal account for half of its citations (see for example (Seglen 1997)). Thus, as emphasized by Garfield (Garfield 2000), "the use of impact factor to weight influence of a paper amounts to a prediction". The ISI founder stressed that impact factors "should not be used as surrogates except in unusual circumstances" (Garfield 1996). It is not practical to exploit actual citation counts, however, when appraising recently published papers. This is simply because it takes several years for many articles to be quoted. Only a small pro-

*This article has not been presented at the conference "The Past, Present, and Future of the Impact Factor and Other Tools of Scientometrics" held in Warsaw (September 26, 2008).

portion of articles (up to 25% in rapidly developing fields such as molecular biology and biochemistry (Garfield 2003)) will have immediate citations ("impact") over the two-year period analyzed by ISI in the IF calculation.

An approach comparing journal performance across subject categories was proposed by Pudovkin and Garfield (Pudovkin and Garfield 2004) called the rank-normalized impact factor. The two-stage analysis presented herein refers to a similar normalization procedure, but in quartile instead of the refined percentile terms used by Pudovkin and Garfield, i.e. the first set studied includes the top 25% of IF-ranked journals (cuts off the 75th percentile) in the subject category and the second group encompasses the bottom 25% of the journal ranking (cuts off the 25th percentile). This bibliometric practice is a starting point for the periodic rating of journals (and, indirectly, papers for the last four years) into four generalized quality classes, a system of so-called "parametric evaluation" performed by the Ministry of Science and Higher Education in Poland for a score-based categorization of research institutions.

Article citedness is seen as fairly predictable (Garfield 1991): "If it's published in a high-impact journal, it is highly likely to be cited. If it is published in a lower-impact periodical, it may remain uncited – even if it received high marks in prepublication peer review or is frequently read". The reliability of this principal bibliometric guess (the "free ride" hypothesis (Seglen 1997)) is examined below because, in fact, empirical literature on such an application of the IF is scarce (see Holden et al. 2006).

A LOW RISK OF BEING UNCITED IN HIGH-IF JOURNALS?

A simple test is based on an estimate of the percent participation of uncited papers from the highest- and lowest-ranked international journals in the quartile IF--based subdivision (Table 1). This bibliometric assessment refers to the journal IF 1998 rating of 59 journals representing the two small and partly overlapping fields geology and paleontology in the Thomson Reuters citation databases, in which article sets from 1996–1997 are reviewed in the context of their citation counts to the end of October 2008. This statistic confirms a negligible likelihood of being overlooked by the scholarly society when results are published in high-IF periodicals (Fig. 1). The mean non-citation rate for the 15 top journals (average IF: 1.8) is a mere 0.88%, and in fact all articles from 9 of them were cited in the more than 10 years after they were published (see also Table 1 in Garfield 1983). In contrast, among 13 low-impact journals (average IF: 0.272), over 22% of the articles (in mean terms) received no citations at all during the same time span; in only one periodical were all papers cited, whilst in two radical cases this lack of citation exceeded 40%.

On the other hand, the guarantee of getting a high citation rate is tremendously increased in well-known journals (mean of the top three most-cited articles: 115 citations, maximum: 299) and contrasts with the highest citation levels in the least prestigious category (average: ca. 22 citations, maximum: 81). A minor overlap of these citation data occurs, but their striking difference is unquestionable, even in these specialties, which are not developing very dynamically at present.

AN INCREASED PROBABILITY OF CITATION SUCCESS IN HIGH-IF JOURNALS?

In a more advanced approach, I analyzed the relationship between the journal IF-based quality ratings from 1992–1994 and real cumulative citation counts to the end of 2001 of Polish geological papers published in 1989-1994 included in the ISI statistical database National Citation Report – Poland 2000 (Racki 2002). Among the 378 articles (citation median: 3), only 79 had a more significant influence because they obtained more than 10 citations (maximum: 92). On the other end, 67 of them were totally ignored (uncited) during the subsequent 7 to 12 years ("failures", arrowed in Fig. 2). This is a frequent case in the geosciences, one of the smaller disciplines covered by the ISI Thomson Reuters databases.

The results are a rather firm test of the hypothetical funding decision in 1995 for the differently evaluated article sets (in terms of arbitrary "parameter points": 24–30 for the high-IF journals and 10 for the low-IF journals according to the system currently used by the Ministry of Science and Higher Education in Poland; see http://www.nauka.gov.pl/mn/_gAllery/40/58/40587//20080826_ujednolicony_wykaz.pdf). Two findings are noteworthy:

- There is a contrast between the citation counts of the papers published in journals exhibiting extremely different IF figures (Fig. 2). The 1st category article set, appearing in the highest-impact quartile of journals (IFs mostly above 1.2, but together with Science and Nature), reveals a median citation value of 5.5. Conversely, those appearing in the low-rated, 4th category journals (IFs mostly below 0.4) commonly received merely 1 citation. The intermediate journal classes exhibit transitional, but very close, values and in fact the 2nd and 3rd (median) IF-based quartiles should be merged.
- Considering the potential inaccuracy of this quality assessment, it is evident that a mere 5.7% of the articles from the least prestigious journals were cited more than 10 times, i.e. these works would be distinctly undervalued in the categorizing and funding level for the authors' institutions by the governmental doner. On the other hand, 12.5% of the articles of the allegedly significant Polish papers from the highest-rated journals were never cited (a noticeably higher non-citation rate than in international publications, shown above, even if in other time spans and

Table 1. List of high- and low-IF journals, based on rankings by IF 1998 from Journal Citation Reports from the ISI subject categories Geology and Paleontology, of which 1996–1997 articles are used to test citedness to the end of October 2008 and the analytical citation data presented in Fig. 1.

Title	IF 1998	Cumulative citations counts of three topmost ("hot") papers	No. of uncited articles*/total no. of articles
Paleontology (26 journals)			
1st category (highest IF 1998 quartile)			
Paleoceanography	3.009	299, 174, 160	0/106
Paleobiology	2.106	103, 89, 78	0/67
Transactions of the Royal Society of Edinburgh – Earth Sciences	2.019	196, 98, 73	0/49
Palaios	1.506	86, 54, 51	0/81
Marine Micropaleontology	1.296	192, 107, 98	4/142
Journal of Quaternary Science	1.250	198, 62, 59	2/69
Palaeogeography Palaeoclimatology Palaeoecology	1.150	99, 98, 95	5/298
4th category (lowest IF 1998 quartile)			
Facies	0.493	50, 42, 40	4/63
Geobios	0.420	21, 19, 18	14/113
Micropaleontology	0.273	81, 55, 46	6/48
Acta Palaeontologica Polonica	0.270	25, 25, 17	0/31
Journal of Micropalaeontology	0.220	25, 21, 20	8/58
Stratigraphy and Geological Correlation	0.056	9, 6, 6	75/127
Paleontologicheskii Zhurnal	0.023	7, 7, 7	88/188
Geology (33 journals)			
1st category (highest IF 1998 quartile)			
Basin Research	2.700	90, 77, 70	0 /32
Journal of Metamorphic Geology	2.188	186, 174, 145	0/106
Geology	2.069	266, 176, 140**	0/273**
Ore Geology Reviews	1.838	61, 58, 12	0/9
Journal of Geology	1.822	117, 105, 100	2/96
Sedimentology	1.624	142, 82, 81	0/109
Palaios	1.506	86, 54, 51	0/81
Journal of Sedimentary Research	1.298	112, 105, 103	3/200
Quaternary Science Reviews	1.220	116, 98, 89	2/138
4th category (lowest if 1998 quartile)			
Newsletters on Stratigraphy	0.500	38, 22, 19	7/34
Proceedings of the Geologists Association	0.500	19, 14, 12	9/54
Facies	0.493	50, 42, 40	4/63
Carbonates and Evaporites	0.367	15, 14, 13	6/37
Exploration and Mining Geology	0.344	46, 14, 12	4/29
Geologie en Mijnbouw	0.273	19, 16, 16	12/56
Stratigraphy and Geological Correlation	0.056	9, 6, 6	75/127
Geology of Ore Deposits	0.035	14, 10, 10	39/140
Revista Geologica de Chile	0.034	37, 25, 19	10/50

*Uncited notes, letters, etc. excluded.
**Only 1996 articles are shown because the large number of articles exceeds the sorting ability of results by times cited in the Science Citation Index Extended.

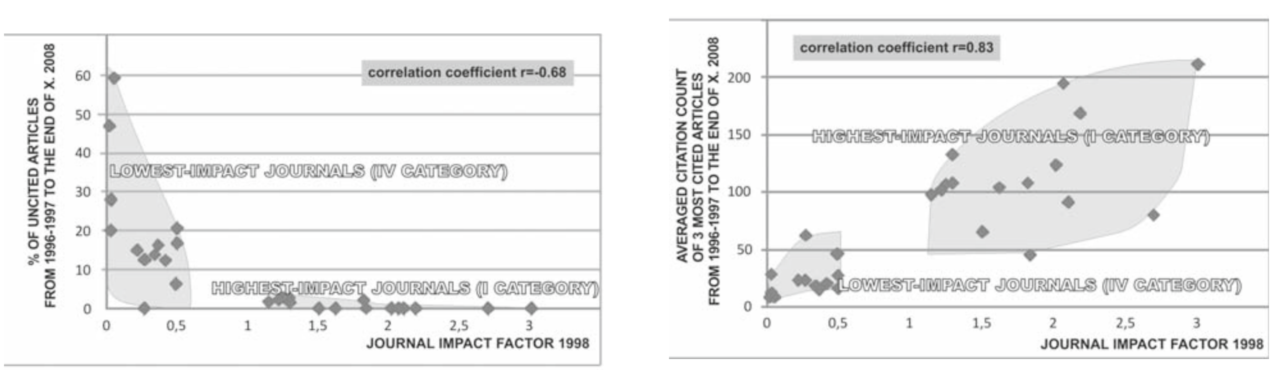

Fig. 1. Uncitedness levels and citation maxima of 1996–1997 articles to the end of October 2008 for the two extreme journal quartiles based on the ranking by IF 1998 for the ISI subject categories Geology and Paleontology (see Table 1).

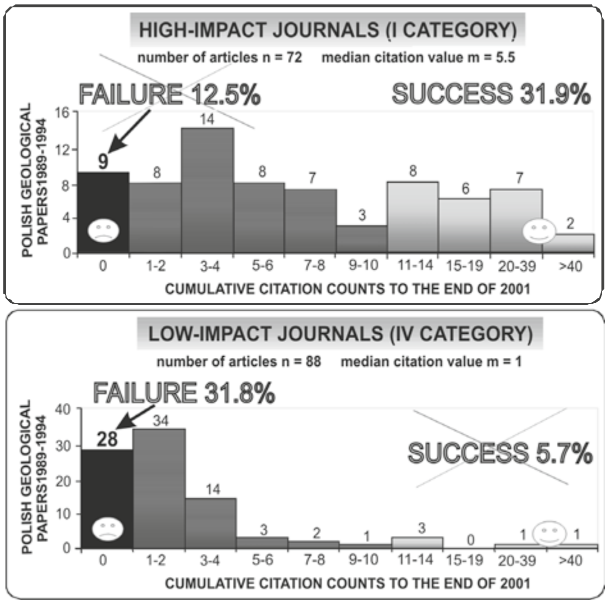

Fig. 2. An assessment of citation success (>10 citations) vs. failure (non-citation, arrowed) to the end of 2001 for Polish geological papers from 1989–1994 published in two extreme journal categories corresponding to the quartile IF 1992–1994-based journal ratings. Crosses show the most severe errors in the use of the IF--based score system for an assumed funding choice, i.e. as a predictive tool of the article citation rate, approximated by the quality of articles.

smaller periodical groups). In more general terms, two thirds of the articles from the high-IF journals achieved a better than average impact (cited more than 4 times), whilst a slightly larger proportion (72%) was more or less overlooked (two or less citations) when published in low-impact periodicals.

Reading these results literally, the probability of non-citation is lessened by a factor of ca. 2.5 in the most highly qualified journals and the likelihood of a relatively high citation rate is increased almost 6 times in those renowned periodicals.

DISCUSSION AND CONCLUSION

The established error levels in the prognosis of expected success versus failure in the scientific "competition for attention" (Franck 1999) with the use of rank--normalized IF classes as an evaluation tool for expected citation numbers are promisingly low (at most 12.5%) in light of the preliminary statistical examination of the 378 articles presented above. This suggestion contrasts more or less with empirical data on a low predictor potential of absolute IF values because implicit application of the indicator to infer the actual citations of individual papers may be largely unreliable (Fava and Ottolini 2000; Filion and Pless 2008; Holden et al. 2006; Leimu and Koricheva 2005; Seglen 1994; Seglen 1997 and references quoted therein). This uncertainty is also confirmed in this study (the correlation coefficient is below 0.5).

Garfield (Garfield 1983; Garfield 1996; Garfield 2006) stressed the fact that each paper accepted in most reputable in-depth peer-reviewed journals, despite a strong author-competitive pressure, means that the scientist met an international quality standard (even if the citation level of the individual article remains uncertain). In this context, the outgoing editor-in-chief of a prominent geological journal, J. Fairchild (Fairchild 2000), remarked: "when looking back at the papers I handled in the early 1990s, the least cited papers have indeed turned out to be those that I was in two minds whether to accept, despite positive referees reports". High impact simply records the ability of the journal's title and editors to gather the best manuscripts available on the scientific market. As recapitulated by Hoeffel (Hoeffel 1998): "Experience has shown that in each specialty the best journals are those in which it is most difficult to have an article accepted, and these are the journals that have a high impact factor (...). The use of impact factor as a measure of quality is widespread because it fits well with the opinion we have in each field of the best journals in our specialty" (see also Kabala 1998; Saha et al. 2003).

On the other hand, journal publishers and editors have become recently obsessed with impact factor scores and there are recognized, mostly unethical, manipulation practices to artificially raise the IF value (see a summary of this "game" in Falagas and Alexiou 2008). Excessive self-citation resulted in an inflation of citation count and IF (Swart and Carling 2008), but external citation criteria are also an intricate issue (see for example (Leimu and Koricheva 2005)). In addition, the relationship between the short-term (impact) and long-term citation frequency obviously is not absolute because an article's, and consequently the journal's, rating can certainly evolve in the course of new discoveries, methodological approaches, improvement in laboratory techniques, and the like. Thus the errors discovered in the predictive validity of rank-normalized IF, up to 12.5% as shown above, are explainable by this collective "information noise". However, this is still an unimpressive number compared with potential pitfalls in the course of more intuitive expert judgments (see for example (Cole et al. 1981; Garfield 1983; Rothwell and Martyn 2000)).

A careful and pertinent use of the rank-normalized journal IF as a statistical proxy of real citation counts and, accordingly, as a predictive tool in *a priori* qualification of publications is a reasonable time- and cost-saving alternative (or in any case a significant supplement) to traditional procedures such as informed peer review (Garfield 1983; Pudovkin and Garfield 2004). It certainly helps overcome the widespread "publish or perish" syndrome. The objective metrics, especially when generalized in terms of rank-normalized journal quality classes across particular fields (Pudovkin and Garfield 2004), may therefore be reasonably applied to an approximate, in probability terms, integrated evaluation of current (only!) research output. Blanket criticism of applying IF in decisions on research funding is therefore at least partly exaggerated. The evaluative practice "merits significant concern as well as further empirical investigation" (Holden et al. 2006) just because "impact factor is not a perfect tool to measure the quality of articles but there is nothing better" (Hoeffel 1998). This IF-based approach is "like the market economy: the system is wrong and unjust, but other systems are much worse" (Łomnicki 2003), which is especially important for scientific progress in developing countries because of the global effect of applying this rating procedure (Łomnicki 2003; Pudovkin and Garfield 2004).

Acknowledgment: I thank the anonymous journal reviewer for constructive suggestions on an earlier version of the manuscript and Professors Norm Savage and Gary Holden and Dr. Marek Wroński for helpful remarks and linguistic adjustment.

REFERENCES

Adam D (2002) The counting house. Nature 415:726–729

Cole S, Cole JR and Simon GA (1981) Chance and consensus in peer review. Science 214:881–886

Fairchild J (2000) Celebration and consideration of citations. J Geol Soc 157:1089–1091

Falagas ME, Alexiou VG (2008) The top-ten in journal impact factor manipulation. Arch Immunol Ther Exp 56:223–226

Fava GA, Ottolini F (2000) Impact factors versus actual citations. Psychother Psychosom 69:285–286

Filion KB, Pless IB (2008) Factors related to the frequency of citation of epidemiologic publications. Epidemiol Perspect Innov Doi:10.1186/1742-5573-5-3 (available via http://www.epi-perspectives.com/ content/5/1/3)

Franck G (1999) Scientific communication – a vanity fair? Science 286:53, 55

Garfield E (1983) How to use citation analysis for faculty evaluations, and when is relevant? Parts 1 and 2. Essays Inf Scientist 6:354–372 (available via http://www.garfield.library.upenn.edu/essays/v6p354y1983.pdf; http://www.garfield.library.upenn.edu/essays/v6p363y1983.pdf)

Garfield E (1991) To be an uncited scientist is no cause for shame. Essays Inf Scientist 14:390–391 (available via http://www.garfield.library.upenn.edu/essays/v14p390y1991.pdf)

Garfield E (1996) How can impact factors be improved? BMJ 313:411–413

Garfield E (2000) Use of Journal Citation Reports and Journal Performance Indicators in measuring short and long term journal impact. Croatian Med J 41:368–374 (available via http://www.garfield.library.upenn.edu/papers/croatianmedj41(4)p368-374y2000.pdf)

Garfield E (2003) The meaning of the Impact Factor. Rev Int Psicol Clinica Salud 3:363–369 (available via http://garfield.library.upenn.edu/papers/meaningofif2003.pdf)

Garfield E (2006) The history and meaning of the Journal Impact Factor. JAMA 293:90–93

Hoeffel C (1998) Journal impact factors. Allergy 53:1225

Holden G, Rosenberg G, Barker K et al (2006) An assessment of the predictive validity of Impact Factor scores: Implications for academic employment decisions in social work. Res Social Work Pract 16:613–624

Kabala ZJ (1998) Know thy journals (available via http://www.agu.org/eos_elec/98073e.html)

Leimu R, Koricheva J (2005) What determines the citation frequency of ecological papers? Trends Ecol Evol 20:28–32

Łomnicki A (2003) Impact factors reward and promote excellence – the system is unkind but effective. Others would do less good for developing countries. Nature 424:487

Pudovkin AI, Garfield E (2004) Rank-normalized Impact Factor: A way to compare journal performance across subject categories. Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology 041:507–515

Racki G (2002) Parametryczny system oceny jednostek naukowych przez KBN: prognozy i postulaty. Zagadnienia Naukoznawstwa 38:51–58

Rothwell PM, Martyn CN (2000) Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? Brain 123:1964–1969

Saha S, Saint S, Christakis D (2003) Impact Factor: a valid measure of journal quality? J Med Libr Assoc 91:42–46

Seglen PO (1994) Causal relationship between article citedness and journal impact. J Am Soc Inf Sci 45:1–11

Seglen PO (1997) Why the Impact Factor of journals should not be used for evaluating research? BMJ 124:498–50225

Swart P, Carling P (2008) Citations and other musings. Sedimentology 55:1115–1116

Wróblewski AK (2008) A commentary on misuses of the impact factor. Arch Immunol Ther Exp 56:355–356

Żylicz M (1999) Biologia molekularna w Polsce. Sprawy Nauki 2:15–18