Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 79, No. 3 (2009) 635–649
DOI: 10.1007/s11192-007-2036-x

History of the journal impact factor: Contingencies and consequences

ÉRIC ARCHAMBAULT,[a,b] VINCENT LARIVIÈRE[b,c]

[a] *Science-Metrix, Montréal, 1335 A avenue du Mont-Royal E, Montréal, Québec H2J 1Y6, Canada*
[b] *Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal, Québec, Canada*
[c] *School of Information Studies, McGill University, Montréal, Québec, Canada*

This paper examines the genesis of journal impact measures and how their evolution culminated in the journal impact factor (JIF) produced by the Institute for Scientific Information. The paper shows how the various building blocks of the dominant JIF (published in the Journal Citation Report - JCR) came into being. The paper argues that these building blocks were all constructed fairly arbitrarily or for different purposes than those that govern the contemporary use of the JIF. The results are a faulty method, widely open to manipulation by journal editors and misuse by uncritical parties. The discussion examines some solution offered to the bibliometrics and scientific communities considering the wide use of this indicator at present.

Introduction

Over the last three decades, librarians and bibliometricians have progressively come to rely on the journal impact factor (JIF). Moreover, interest in this indicator and its derivatives has grown exponentially in the scientific community since 1995 (Figure 1). Many researchers have observed that the indicator is driving the publishing strategies of scientists who want to maximize their average impact factor and how, similarly, journal editors aspire to augment their JIF by using strategies that sometimes diverge considerably from widely held beliefs on the basic ethics of science (see, e.g., [SMITH, 1997]). Moreover, it is not uncommon to find these indicators being used to promote researchers (see, e.g., [FUYONO & CYRANOSKI, 2006]. In response, bibliometricians have increasingly tried to "tame the beast" by suggesting numerous improvements aimed at increasing the validity of the JIF as a quantitative measure. Despite this growing interest, there is, apart from GARFIELD's historical accounts [2006] and intellectual biographies [BENSMAN, 2007], a notable scarcity of contributions to the conceptual history of this important indicator.

Received June 19, 2008; Published online January 31, 2009

Address for correspondence:
ÉRIC ARCHAMBAULT
E-mail: eric.archambault@science-metrix.com

0138–9130/US $ 20.00

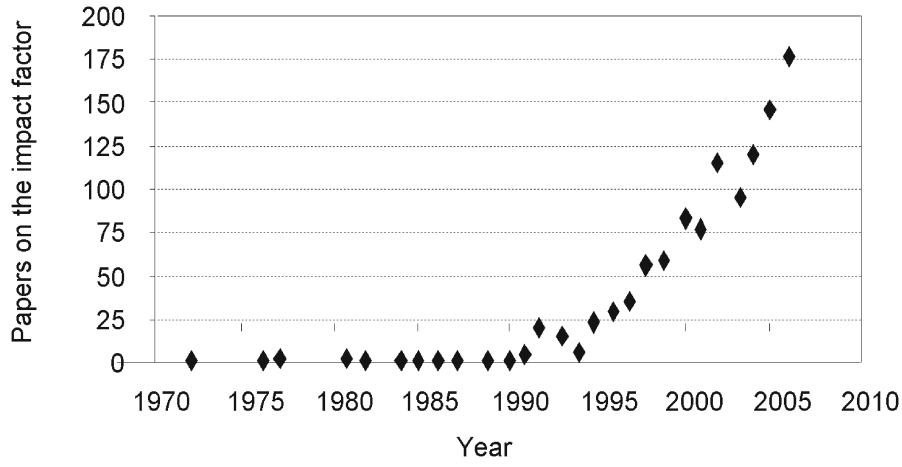

Figure 1. Number of papers on the impact factor indexed in Thomson Scientific's Web of Science, 1963–2006

This paper provides a historical account of the piecemeal development of methods that culminated in the production of the JIF. It would be beyond the purposes of the paper to attempt to capture every characteristic of this indicator, its origins, and its effects on bibliometrics and science evaluation. Rather, the paper concentrates on five salient characteristics of the JIF and how they emerged and how these sometimes accidental characteristics shaped their subsequent use in ways that were surely unforeseen by the empiricists who contributed to what was ultimately presented as a coherent system for the measurement of scientific journals' impact.

Firstly, the tool's origin and development was guided by the needs of US university and college librarians who wanted to use an objective method to select journals for their holdings. This had at least two consequences: the tool was not initially developed for research evaluation, and the approach was clearly optimized for the US context. Secondly, an important consideration is that, early on, methods allowed for the identification of high impact journals in specific fields, and this subsequently raised issues of misuse in research evaluation and outright abuse in the promotion of researchers. Thirdly, a central aspect of the methods developed to determine journal impact measures is the consideration given to journal self-citations, which did not pose a particular problem until journal editors started to encourage authors to consider citing papers published in the journal where they had submitted their article for publication. Fourthly, another problem is the asymmetry between the document types that are counted in the numerator and in the denominator of the JIF equation, which again did not cause problems until journal editors realized that they could fairly easily modify their

journals' ratings by reengineering the types of documents being published. Finally, based on convenience rather than on empirical reasons, it was decided to use a two-year citation window, which causes problems due to varying citation latencies between article types, journals, and fields; again, this opened the door to manipulation by journal editors.

This paper examines how these features came about and discusses the impact of these characteristics considering the *actual* use, as opposed to the hypothetical or officially prescribed use, of this metric. The discussion examines the various solutions available to the bibliometrics and scientific community.

A method developed by US university librarians for US university librarians

The literature on the use of journal impact measures uniformly concludes that GROSS & GROSS [1927] were the first to develop this method (see, e.g., [ALLEN, 1929; MCNEELY & CROSNO, 1930; GROSS & WOODFORD, 1931; HENKLE, 1938; BRODMAN, 1944; GARFIELD, 1955; RAISIG, 1960]. Gross and Gross sought to address the rising problems of small colleges at a time when one "of the biggest of these [was] the problem of adequate library facility." GROSS & GROSS [1927] raised a question that is still highly relevant today: "What files of scientific periodicals are needed in a college library successfully to prepare the student for advanced work, taking into consideration also those materials necessary for the stimulation and intellectual development of the faculty?"

Gross and Gross rhetorically considered the compilation of a list of relevant journals using a subjective approach; this strategy intended to outline the advantages of their more objective method: "One way to answer this question would be merely to sit down and compile a list of those journals which one considers indispensable. Such a procedure might prove eminently successful in certain cases, but it seems reasonably certain that often the result would be seasoned too much by the needs, likes and dislikes of the compiler."

Thus, one can note that the first use of journal impact calculation aimed to facilitate the task of journal selection using objective quantitative methods, which is one core aspect in the marketing of the most visible commercial product that has emerged from this work – Thomson Scientific's (formerly the Institute for Scientific Information, or ISI) Journal Citation Report (JCR). Moreover, an important feature in the development of this method was that at the outset it was developed specifically to cater to the needs of US librarians. For example, in a study of mechanical engineering, MCNEELY & CROSNO [1930] stated: "It will be noted that the list contains three American, one English, two German publications, and one French publication. The result is that the English language publications predominate, but it is assumed that such should be the case for American libraries."

What is obvious when one examines the evolution of journal impact calculations that followed Gross and Gross' seminal paper is the growing complexity and size of the compilations (Table 1). The paper by GROSS & GROSS [1927] proved to be an inspiration for several US librarians and early information scientists. For instance, BRODMAN [1944] cited no less than 18 papers published after 1926 that used a method based on the Gross and Gross paper – all of them from the US.

Table 1. Growth in the number of references and sources

Authors	Sources	References	# of fields	Field
Gross and Gross, 1927	1	3633	1	Chemistry
Allen, 1929	9	2165	1	Maths
McNeely and Crosno, 1930	7	17991	1	Electrical Engineering
Gross and Woodford, 1931	6	3574	1	Geology
Gregory, 1934	~40	26760	1 (27 sub.)	Medical
Hooker, 1935	5	12794	1	Physics
Hackh, 1936	20	22575	1	Dentistry
Henkle, 1938	1	17198	1	Biochemistry
Brown, 1956	57	~38000	8	Several

The emergence and evolution of this method on US soil, particularly after the spectacular scale up made by ISI in the 1970s, likely had the effect of creating a self-fulfilling prophecy. Indeed, by concentrating on the US situation and by positively biasing the sources in favour of US journals, the method placed these journals on centre stage. Had a broader linguistic and national coverage been considered, it might have revealed that these journals were not in fact more cited than others. By creating this centre stage, the measures of JIF made a selective promotion of US journals, which could then be picked up, read, and increasingly cited by researchers in the US and also abroad. This positive feedback process certainly shaped and self-reinforced the importance of US-dominated journals over the 30 years or so that the JIF has been available. Had the Institute for Scientific Information emerged as the "Institut für Forschungsinformation", the JCR would undoubtedly have evolved in a substantially different form and, the aggregate current impact of German journals would likely be substantially larger.

A method centred on scientific fields

In the early years of the method's development, studies were generally limited to a single field (e.g., chemistry, geology, or medicine), and references were often compiled from a single key journal (e.g., [ALLEN, 1929]) or key reference monograph (e.g., [HACKH, 1936]). These papers essentially copied Gross and Gross's selected approach, which, in their case, involved compiling citations from a key source in the field of chemistry – the 1926 edition of the *Journal of the American Chemical Society*. This resulted in a compilation comprising 3,633 references to 247 journals. By 1930, the

method had gained two additional characteristics: several journals were used as sources and, although the practice was still predominantly centred on the English language, many non-US source journals were included. For example, MCNEELY & CROSNO [1930] used seven source journals and compiled a total of 17,991 references. In a similar manner, GREGORY [1937] produced a colossal study – considering the technical means available at the time – using the Gross and Gross method to identify key journals in 27 subfields relevant to medicine, in which they tabulated some 26,700 references from about 40 source journals or monographs. BROWN [1956] published a monograph entitled *Scientific Serials,* basing the approach, by then generalized, on collecting citations from several journals. Brown covered eight fields of science using 57 source journals and a compilation of close to 38,000 references. Later, MARTYN & GILCHRIST [1968] produced a study that resembled, in both shape and scope, the current JCR, with its wide variety of fields.

Importantly, all of the aforementioned studies produced field-specific listings, and no author saw a need to adapt the method to enable comparisons between fields. Even authors such as GREGORY [1937], BROWN [1956], and MARTYN & GILCHRIST [1968] who produced measures for several fields or specialties always presented scores separately for each specialty. There was, in fact, no need for cross-field comparability, as the purpose of the technique was to identify relevant journals for libraries to have adequate journal coverage of a series of distinct fields.

However, this practice produced adverse effects when measures of journal impact were used to evaluate scientific production across fields. One of the most blatant abuses of the commercial tools that have emerged out of these methods involves giving bonuses or making promotion decisions for researchers based on raw impact factor values that do not take account of the fact that these measures are field-specific. For example, FUYONO & CYRANOSKI [2006] mention that, in Pakistan, researchers can earn bonuses amounting to anywhere between $1,000 and $20,000 based on the cumulative one-year impact factor of the journals in which they publish. The authors also provide the example of the Chinese Academy of Sciences' Institute of Biophysics, which has a scale tuned to the impact factor: publications in journals with a JIF between 3 and 5 are worth 2,000 yuan per JIF point, and a publication in a journal with a score higher than 10 is worth 7,000 yuan per JIF point. For anyone who has worked with or read about the impact factor, it is well-known that JIF scores vary tremendously between fields. To careful users of the JIF, it becomes clear that schemes such as these are helping researchers in the biomedical field become wealthier (because these fields have high citation rates and therefore high non-normalized impact factor values), while others, such as mathematicians or social scientists, are obtaining only small bonuses (because of the lower citation propensity in their fields), even if they manage to publish in the best journals of their respective fields.

The Institute for Scientific Information has been aware of the problem of inter-field comparability for a long time. For example, Garfield suggested that:

> Instead of directly comparing the citation count of, say, a mathematician against that of a biochemist, both should be ranked with their peers, and the comparison should be made between rankings. Using this method, a mathematician who ranked in the 70 percentile group of mathematicians would have an edge over a biochemist who ranked in the 40 percentile group of biochemists, even if the biochemist's citation count was higher.[1] [GARFIELD, 1979, as cited in SCHUBERT & BRAUN, 1996, P. 312]

Despite this knowledge, ISI has continued for nearly 30 years to publish the JCR with data that is not comparable across fields, thus leaving the door open to the abuse of this increasingly popular metric.

The inclusion of journal self-citations

The debate over whether to include journal self-citations is as old as the methods to calculate measures of journal impact. GROSS & GROSS [1927] decided to exclude journal self-citation. This might have been because they saw the huge impact of journal self-citation, or it may have been simply because they felt that every library should own the source used in their seminal study. This practice, and the important impact of journal self-citations, was discussed by ALLEN [1929], MCNEELY & CROSNO [1930], GREGORY [1937], and HOOKER [1935], but all of these authors made the decision to exclude self-citations. WESTBROOK [1960], who used citations to evaluate research as opposed to studying the impact of scientific journals, decided to exclude "in-house" citations, though he claimed that this was perhaps unnecessary. GARFIELD & SHER [1963B] distanced themselves from Westbrook and mentioned that (journal) self-citations should be included in a JIF.

Unfortunately, Garfield and Sher did not foresee that their choice provided a potent way for editors to manipulate the JIF by inducing authors to cite the journal in which they publish; because journal self-citations are counted, it is also possible to influence the JIF by "encouraging" authors to cite papers from the journal in which they *seek* to be published. For instance, this letter was sent to authors wishing to publish in *Leukemia* [SMITH, 1997]:

> Manuscripts that have been published in *Leukemia* are too frequently ignored in the reference list of newly submitted manuscripts, even though they may be extremely relevant. As we all know, the scientific community can suffer from selective

[1] Following on this idea, PUDOVKIN & GARFIELD [2004] suggested using a rank-normalized impact factor using percentiles.

> memory when giving credit to colleagues. While we have little power over other journals, we can at least start by giving you and others proper credit in *Leukemia*. We have noticed that you cite *Leukemia* [once in 42 references]. Consequently, we kindly ask you to add references of articles published in *Leukemia* to your present article (p. 463).

The emergence of the asymmetry between numerator and denominator

HACKH [1936] proposed the idea of dividing the number of references by the number of volumes, thus, for the first time, taking into account the extent of the citable material. This idea was not taken up in the literature until 1960, when it reappeared in the work of RAISIG [1960]. By and large, the approach suggested by Raisig involved taking into consideration the "relationship of the number of articles quoted to the number of articles published," a method which was coined as the RPR index or "index of research potential realized" [RAISIG, 1960, P. 1418]. Raisig's suggestion to use a ratio of citations to source articles was subsequently espoused by GARFIELD & SHER [1963B] for the calculation of a would-be "journal impact factor".A few years later (but possibly without prior knowledge of Raisig's approach), MARTYN & GILCHRIST [1968] also decided to exclude some source items, such as abstracts, obituaries, reviews, and bibliographies, from the counts. However, in contrast to Raisig, who went to great lengths to symmetrically count citations made to "original articles" and the corresponding number of original articles by using the data contained in the SCI, and given the limited power of computers at the time, it was not possible (or at the very least there were sizeable difficulties) for Martyn and Gilchrist to associate the great number of citations to the large number of source items contained in the SCI data they used. Today, there are algorithms that routinely do just that, and there are no technical reasons for having an asymmetrical count in the calculation of citation and source items – only path dependency and a technical look-in can explain this important shortcoming.

Clearly, the asymmetry between what is counted in the numerator (references to every type of material) and what is counted in the denominator (only the types of documents that are deemed citable) predates ISI's impact factor. The fact that Raisig, as well as Martyn and Gilchrist, were cited by GARFIELD [1972], a turning point in the development of the ISI impact factor, strongly suggests that this characteristic was adopted from the prior art rather than invented by Garfield and his colleagues at ISI.

Of the technical limits to the dominant method for journal impact measurement, the asymmetry between the numerator and denominator and journal self-citations is among the most commonly mentioned. This asymmetry induces some strong distortions, more particularly for highly cited journals [MOED & VAN LEEUWEN, 1995]. This can also lead to manipulation on the part of editors who multiply source items that are not considered "citable" but are in fact cited frequently. As a matter of fact, in several

fields, the journals with the highest JIF are review journals, simply because review articles are often more frequently cited than regular articles. It is a widely held secret that this presents an opportunity for shrewd editors to distort JIF scores by simply increasing the number of reviews in their journals.

A convenient, but arbitrary, citation window

Another important attribute of ISI's impact factor is the controversial two-year citation window that was developed by MARTYN & GILCHRIST [1968]. It is worth quoting the authors at length to show that although they had thought of possible consequences, this citation window was chosen for its convenience, considering the data they had in hand:

> We decided that our most practical course would be to confine our study to citations made during 1965 to journals published in the two preceding years. It was already known that 26.1% of all the 1965 citations were to literature of 1964 and 1963, so in terms of number of citations this would give us an adequate sample. There is reason to suppose that, so far as the more important journals are concerned, the ranking we obtained would have been materially altered had our sample covered a greater time span, and by confining ourselves to the two years prior to 1965, we avoided the problem of correcting for cited age […]. Taking the two-year sample reduced the effort of counting and also reduced the cost of acquisition of the data (p. 2).

Five years later, GARFIELD [1972: 476] mentioned his indebtedness to Martyn and Gilchrist's method:

> To calculate an impact factor for each journal, I divided the number of times 1967 and 1968 articles were cited in 1969 by the number of articles published in 1967 and 1968. Martyn and Gilchrist used a similar method in ranking British journals in an analysis of 1965 SCI data.

As this is the method still used today to calculate Thomson Scientific's JIF, one can surmise that the use of this approach results from a choice to follow the most practical course, one that reduced the effort of counting and the cost of acquiring data rather than one that was the result of an in-depth analysis of various solutions and the subsequent choice of optimal characteristics. Indeed, Garfield was well aware that the vast majority of citations were older than two years: not only was it mentioned in Martyn and Gilchrist, but in 1963, Garfield noted that "over 50% of the cited references in the 1961 index are more than five years old" [GARFIELD & SHER, 1963A]. This statement indicated that Garfield was aware that the half-life of the cited references was greater

than five years and that going back two years certainly meant missing out on a very substantial part of the impact picture.

The two-year citation window used in the JCR is repeatedly criticized in the bibliometrics community. There is a consensus that the calculation of the impact factor based on a citation window of only two years is far too short in many fields. More importantly, it is now known that one basic assumption made by Martyn and Gilchrist ("There is reason to suppose that, so far as the more important journals are concerned, the ranking we obtained would have been materially altered had our sample covered a greater time span") was made in error. For instance, GLÄNZEL & MOED [2002] cite the example of the comparison between the impact of *The Lancet* and the *American Sociological Review* (ASR). When a short citation window is used, *The Lancet* has a greater mean citation rate, but when using a window of four years or more, it is the ASR that has a higher mean citation rate.

This is yet another example suggesting either that the basic building block of the JIF was based on an un-proven formula or that it was optimized to meet the needs of US university librarians who had very specific needs at the time that they developed and used the various elements that made their way into the JCR.

Discussion

The evidence presented here strongly suggests that Martyn and Gilchrist were the actual creators of the impact factor as we know it. GARFIELD [1972] adopted their method, and ISI's JCR essentially provided a massive scale-up in the use of existing techniques, permitted by ISI's ownership of the construction of the Science Citation Index. Like most of the pioneers in the calculation of journal impact measures, whose work was aimed at fulfilling the needs of US scientific librarians rather than unambiguously determining the best journals regardless of country of origin and language, Garfield's company used source journals that were undoubtedly US- and English-language-centric.

Importantly, had this method been developed in a different country and had Garfield been, for instance, German, he would have started by using a majority of German-language journals as source items. This would have resulted in journal impact values that were different from those he obtained and the progressive inclusion of a set of journals that would most likely have resulted in a different source database than the one commercialized today by Thomson Scientific, and which incidentally serves to calculate today's JIF. Had this been the case, the impact factor values presented in the JCR would surely be substantially different. Moreover, it is apparent that there was a self-fulfilling prophecy in the publication of these scores: as the JCR claimed that some journals had more impact than others, scientists undoubtedly gradually turned to these journals for publication and also as sources of information. This has gradually resulted

in these journals being more cited, regardless of whether they were more frequently cited than journals that had a greater objective impact considering the whole of the international scientific system (that is, inclusive of all journals, in all countries, and in all languages). Thus, the JCR has reshaped international science in favour of both the US and the English language. Simulations and modelling would help us to understand the extent of this reshaping, but it is likely measurable in percentage points, which is considerable given that it is the result of the work of a single firm.

For all of the reasons presented in this paper, the indicators presented in the JCR cannot be considered robust objective measures of the worth of all journals published internationally. The JCR and its JIF measures, although the result of historically contingent events and highly subjective choices, have had a profound effect on the way research, journals, and even scientists are evaluated around the world today. There are certainly several objections that could be raised against using measures such as the JCR. For instance, for the sake of convenience, one can categorize objections to the use of journal impact measures into three groups: 1) scientific activities should not be evaluated using bibliometric methods, particularly the impact factor; 2) indicators can be easily misused, but there are relatively simple normalizations that can, in general, prevent this misuse; 3) indicators are technically flawed but can be re-engineered in depth, and their flaws corrected.

The first type of objection arises mostly from epistemological reasoning. Examples of arguments used to unequivocally reject the use of citation analyses and journal impact measures include:

- Some scientific works are only recognized several years after their publication, while any citation analysis is limited to a predetermined citation window [LINDSEY, 1989].
- Papers that are never cited do not necessarily have zero impact [SEGLEN, 1997].
- Negative citations are counted in the same way as positive citations [OPTHOF, 1997].

Although bibliometricians will generally recognize some, if not all, of these limits, they will usually counter such arguments by stating that the strength of their indicators is conferred by the law of large numbers and that this is to the levels of aggregation. GLÄNZEL & MOED [2002] distinguish between three levels of aggregation: 1) the micro level (individual scientist, research group); 2) the meso level (institutions, journals); and 3) the macro level (national and supra-national research, subject analyses). As noted by SEGLEN [1997], "[s]ince any large, random sample of journal articles will correlate well with the corresponding average of the JIF, the impact factor may seem reasonably representative. However, the correlation between journal impact and actual citation rate of articles from individual scientists or research groups is often poor."

For macro-level analyses, several of the weaknesses of the JIF are attenuated, but there is one aspect that nearly everyone who carefully uses these indicators will become aware of: the JIF needs to be modified somewhat to take into account inter-field variations. Moreover, using normalization to even out scores across fields, it is possible to correct one of the most obvious deficiencies of the JCR, which is to provide field-specific measures in an environment where it is used in areas that encompass several fields. Countless papers provide suggestions on how to normalize for differences across fields.[2] For many bibliometricians, this type of correction is mainly useful for performing studies at more macro levels. For instance, SEGLEN [1992] argues that as long as corrections are made to account for differences across fields, "citedness can be a useful indicator of scientific impact at the national level" (p. 637). Even so, some bibliometricians would advise against this type of usage, given the numerous flaws of the currently dominant JIF.

The second type of objection is that measures of journal impact are prone to be manipulated, misused, and even abused. The analysis of the genesis of journal impact measures in this paper made it apparent that these indicators were developed with a clear intent – that is, to support the work of librarians in managing their journal collections. Likewise, Thomson Scientific has explained in simple terms the intended uses of the JCR.[3]

Enables a variety of information professionals to access and assess key journal data:

- *Librarians* can manage and maintain journal collections and budget for subscriptions […].
- *Publishers* can monitor their competitors, identify new publishing opportunities, and make decisions regarding current publications.
- *Editors* can assess the effectiveness of editorial policies and objectives and track the standing of their journals.
- *Authors* can identify journals in which to publish, confirm the status of journals in which they have published, and identify journals relevant to their research.
- *Information Analysts* can track bibliometric trends, study the sociology of scholarly and technical publications, and study citation patterns within and between disciplines.

However, one must not be naive about the prescriptions of Thomson Scientific. Although the official message is that the JCR should not be used to evaluate research output let alone to provide bonuses or promote academics, everyone knows that this is

[2] See e.g. FASSOULAKI & AL. [2002], HUTH [2001], RAMIREZ & AL. [2000], SCHWARTZ & LOPEZ HELLIN [1996], SEN & SHAILENDRA [1992], SOMBATSOMPOP & AL. [2005] and VAN LEEUWEN & MOED [2001] to name but a few.

[3] http://scientific.thomson.com/products/jcr/

the current situation and that Thomson Scientific benefits both economically and in terms of prestige by the very widespread use of the JCR. Yet, for all of the sometimes vehement critiques made towards this indicator, Thomson Scientific has yet to modify this important product to minimize its biases and reduce its vulnerability to manipulation.

Though this paper, like many before, has shown that the JIF used in the JCR is fairly easy to manipulate, it would have been beyond the aim of this paper to show all the possible instances of manipulation, let alone all the ways in which it can be manipulated. Given the current knowledge of the situation, there is a need for journal editors to address the misuse of this indicator in a pro-active manner. Scientific quality is supposed to be driving the scores obtained in measures of journal impact, and this should become and remain the sole focus of their application.

The third type of objection is primarily technically based, such as the inadequacy of two-year citation windows in many cases, or the asymmetry between items counted in the numerator and denominator. Unlike the first type of objection, it was traditionally voiced by bibliometrics researchers but, as can be seen in the editorials of biomedical and clinical research journals, a large number of researchers have primarily negative comments about the limits of this indicator. This last type of objection presents a greater problem because these shortcomings can often be solved only by having access to all of the source data, and very few bibliometricians, not to mention scientists in general, have that kind of access. The few teams that have access to source data can certainly produce corrected measures of impact factors and use these measures for their own research and contractual undertakings, but they would not be authorized to commercialize these indicators and compete against the JCR.

The JCR and its measures of journal impact have some significant shortcomings, but these clearly have different levels of consequence, depending on the use that is made of measures of journal impact. For the uses officially intended by Thomson Scientific, such as selecting journals for a library, the weaknesses are certainly acceptable, although one has to acknowledge that this measure is probably reshaping science at the cost of reduced national and linguistic diversity. When used for policy making at the national level, it becomes important to normalize by field to obtain an adequate picture. As one goes down the scale of applications, it becomes absolutely imperative to normalize data; the deficiencies of the impact factor also become more worrisome, for the laws of large numbers decreasingly come into play to compensate for the shortcomings of the way in which JCR metrics are computed. While some improvements (e.g., field normalization) can be made "in-house" without a large infrastructure, most other improvements can only be made by having access to the source data, which is not the case for most users, especially those outside of the bibliometric community. At the other extreme, the US- and English-language-centeredness may not be correctable through the use of Thomson Scientific databases, so it is not likely that the debate on the limits of these tools will cease.

Some solutions and implications

Given the existing knowledge of the limits of the dominant measure of journal impact, and given that there is every reason to believe that journal impact indicators will continue to be misused and manipulated, it has become increasingly pressing to ask what we should do at this juncture. Considering the types of criticism made of this indicator, we could 1) abandon the use of the JCR for research evaluation and policy; 2) use it and implement measures to ensure that it is used intelligently; or 3) redesign the tool from the ground up, taking into account current awareness and knowledge. Abandoning the use of the JCR, a decision that would be made mainly on epistemological grounds, is very unlikely to happen if current trends are any indication of future practices. Using the indicator intelligently by relying on the law of large numbers to mitigate its imperfections and adapting it for inter-field variations is one road that has already been taken by several "poor-man" bibliometricians who are able to live with imperfections given their limited means. However, there is little chance that non-bibliometricians will be convinced or educated enough to use this metric both wisely and responsibly.

At this juncture, the best course of action therefore seems to be redesigning the tool from the ground up. We could deal knowingly with inter-field variations, journal self-citations, citable and non-citable items, citation windows – all technical aspects for which several teams have already suggested remediation and potentially robust solutions. But we also have to accept that we are faced with a hard combinatorial optimization problem (see [ZIPF, 1949]) – despite what certain authors may write, their measure is not the *correct* measure but merely a *corrected* measure – and therefore deciding on the best compromise for several aspects of the indicator will be a considerable challenge. How best to deal with journal self-citations and the length of the citation window are certainly potential topics for future lively debates. We will also have to consider that there might be more than one design to solve the many problems that exist in highly variable circumstances. This raises the question of how open the market is to a number of new solutions diffused by various parties. It is also necessary to consider the consequences of no longer having and upholding a "gold standard", which has been represented by the JCR to date. Another important question is how we should deal with the "arms race", as we are certain that whatever form the indicator takes, journal editors are likely to respond with equally original ways to manipulate it. In this context, should the formula be kept secret, as far as this is possible, to limit its potential exploitation? Should the behaviour of journal editors be monitored and sanctions be applied (such as excluding journals from such a measure) to prevent that which is considered unacceptable?

But the most important question we should consider is 'who are "we"'? At present, the bibliometric and scientific communities appear to be excluded from this "we", as they have so far been largely unable to effect significant change. The evolution of the dominant indicator and the provision of alternate, widely available solutions have been placed into the hands of a private firm that has failed to respond to increasingly numerous calls for action. The bibliometric and scientific communities can only hope that a response will come from the private sector or that one day, they gain open access to the means necessary to create, share, openly debate, and, hopefully, improve tools to evaluate the impact of journals, so that the resulting indicators could be used sensibly in both research policy and evaluation.

*

The authors wish to thank Jean-Pierre Robitaille, Yves Gingras, the anonymous reviewers, as well as the audience at the ISSI conference in Madrid, for their valuable comments and suggestions. They also wish to thank Julie Caruso and Johanna Kratz for revising the manuscript.

References

ALLEN, E. S. (1929), Periodicals for mathematicians. *Science*, 70 (1825) : 592–594.

BRODMAN, E. (1944), Choosing physiology journals. *Bull Med Libr Assoc*, 32 (4) : 479–483.

BENSMAN, S. J. (2007), Garfield and the impact factor. *Annual Review of Information Science and Technology*, 41 : 93–155.

BROWN, C. H. (1956), Scientific serials: characteristics and lists of most cited publications in mathematics, physics, chemistry, geology, physiology, botany, zoology, and entomology. *ACRL Monograph no. 16*. Chicago: Association of College and Research Libraries.

CUNNINGHAM, E. R. (1935), The present status of the publication of literature in the medical and biological sciences. *Bull Med Libr Assoc*., 24 (1) : 64–81.

FASSOULAKI, A., PAPILAS K., PARASKEVA A., PATRIS K (2002), Impact factor bias and proposed adjustments for its determination. *Acta Anaesthesiologica Scandinavica*, 46 (7) : 902–905.

FUYUNO, I., CYRANOSKI, D. (2006), Cash for papers: Putting a premium on publication. *Nature*, 441 (7095) : 792.

GARFIELD, E. (1955), Citation indexes for science. *Science*, 122 (3159) : 108–111.

GARFIELD, E. (1972), Citation analysis as a tool in journal evaluation. *Science*, 178 (4060) : 471–479.

GARFIELD, E. (2006), The history and meaning of the journal impact factor. *JAMA – Journal of the American Medical Association*, 295 (1) : 90–93.

GARFIELD, E., SHER, I. H. (1963A), *Genetics Citation Index*. Philadelphia: Institute for Scientific Information.

GARFIELD, E., SHER, I. H. (1963B), New factors in evaluation of scientific literature through citation indexing. *American Documentation*, 14 (3) : 195–201.

GLÄNZEL, W., MOED, H. F. (2002), Journal impact measures in bibliometric research. *Scientometrics*, 53 (2) : 171–193.

GREGORY, J. (1937), An evaluation of medical periodicals. *Bull Med Libr Assoc*., 25 (3) : 172–188.

GROSS, P. L. K., GROSS, E. M. (1927), College libraries and chemical education. *Science*, 66 (1713) : 385–389.

GROSS, P. L. K., WOODFORD, A. O. (1931), Serial literature used by American geologists. *Science*, 73 (1903) : 660–664.

HACKH, I. (1936), The periodicals useful in the dental library. *Bull Med Libr Assoc*., 25 (1–2) : 109–112.

HENKLE, H. H. (1938), The periodical literature of biochemistry. *Bull Med Libr Assoc*., 27 (2) : 139–147.

HUTH, E. J. (2001), Authors, editors, policy makers, and the impact factor. *Croatian Medical Journal*, 42 (1) : 14–17.

LINDSEY, D. (1989), Using citation counts as a measure of quality in science: measuring what's measurable rather than what's valid. *Scientometrics*, 15 (3–4) : 189–203.

MARTYN, J., GILCHRIST, A. (1968), *An Evaluation of British Scientific Journals* (1 ed.): Aslib.

MCNEELY, J. K., CROSNO, C. D. (1930), Periodicals for electrical engineers. *Science*, 72 (1856) : 81–84.

MOED, H. F., VAN LEEUWEN, T. N. (1995), Improving the accuracy of Institute for Scientific Informations journal impact factors. *Journal of the American Society for Information Science*, 46 (6) : 461–467.

OPTHOF, T. (1997), Sense and nonsense about the impact factor. *Cardiovascular Research*, 33 (1) : 1–7.

PUDOVKIN, A. I., GARFIELD, E. (2004), Rank-normalized impact factor: A way to compare journal performance across subject categories. *Proceedings of the 67th ASIS&T Annual Meeting*, 41 : 507–515.

RAISIG, L. M. (1960), Mathematical evaluation of the scientific serial. *Science*, 131 (3411) : 1417–1419.

RAMIREZ, A. M., GARCIA, E. O., DEL RIO, J. A. (2000), Renormalized impact factor. *Scientometrics,* 47 (1) : 3–9.

SCHUBERT, A., BRAUN, T. (1996), Cross-field normalization of scientometric indicators. *Scientometrics*, 36 (3) : 311–324.

SCHWARTZ, S., HELLIN, J. L. (1996), Measuring the impact of scientific publications. The case of the biomedical sciences. *Scientometrics*, 35 (1) : 119–132

SEN, B. K., SHAILENDRA, K. (1992), Evaluation of recent scientific research output by a bibliometric method. *Scientometrics*, 23 (1) : 31–46.

SEGLEN, P. O. (1992), The skewness of science. *Journal of the American Society for Information Science*, 43 (9) : 628–638.

SEGLEN, P. O. (1997), Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314 : 497.

SMITH, R. (1997), Journal accused of manipulating impact factor. *British Medical Journal*, 314 (7079) : 463.

SOMBATSOMPOP, N., MARKPIN, T., YOCHAI, W., SAECHIEW, M. (2005), An evaluation of research performance for different subject categories using Impact Factor Point Average (IFPA) index: Thailand case study. *Scientometrics*, 65 (3) : 293–305.

VAN LEEUWEN, T. N., MOED, H. F. (2001), Development and application of new journal impact measures. Cortex, 37 (4) : 607–610.

WESTBROOK, J. H. (1960), Identifying Significant Research. *Science*, 132 (3435) : 1229–1234.

ZIPF, G. K. (1949), *Human Behavior and the Principle of Least Effort*. Cambridge, Ma.: Addison-Wesley.