Origins of Measures of Journal Impact: Historical Contingencies and Their Consequences on Current Use[1]

Éric Archambault* and Vincent Larivière**

*eric.archambault@science-metrix.com
Science-Metrix, 4572 avenue de Lorimier, Montréal, Québec H2H 2B5 (Canada) and
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)

**lariviere.vincent@uqam.ca*
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succursale Centre-ville, Montréal, Québec, H3C 3P8 (Canada) and
Graduate School of Library and Information Studies, McGill University, Montréal, Québec (Canada)

Abstract
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 form of the JIF, which is the result of historically contingent choices rather than a carefully chosen and tested set of features, affected its subsequent use, misuse, and manipulation by researchers, journal editors, and bibliometricians.

Keywords
journal impact factor; history; origins; misuse; use; contingencies.

Introduction

In the last three decades, librarians and bibliometricians have progressively come to rely on the journal impact factor (JIF). Particularly in the late 1990s, the indicator attracted a significant amount of attention in the scientific community. Many researchers have observed that this indicator is orienting the publishing strategies of scientists who want to maximize their impact factor and how, similarly, journal editors aspire to augment their journal's JIF. Consequently, bibliometricians increasingly try to "tame the beast" by suggesting numerous ways to improve the validity of the JIF as a quantitative measure. This growing interest is illustrated by an increase in the number of papers dealing with the indicator, rising from 23 papers indexed in Thomson Scientific's *Web of Science* in 1995 to 146 papers in 2005. Despite this growing interest, there is, apart from Eugene Garfield's own historical accounts, a real scarcity of contributions to the conceptual history of this important indicator. This paper provides an account of the history of the JIF and its subsequent use, misuse, and manipulation by researchers, journal editors, and bibliometricians. It would be beyond the purposes of this paper to attempt to capture every minute characteristic of this indicator as well as its origins and its effects on the evolution of the bibliometric field. Here, we will concentrate on five aspects that have received the largest share of interest: the fact that the indicator was developed to help the management of scientific journal collection, not the evaluation of scientific research; the field-specific nature of the scores produced by the indicator (scores are not readily comparable across scientific disciplines); the asymmetry between what is counted in the numerator and in the denominator; the two-year citation window; and how the scores are English language- and US-centred.

Origins of Measures of Journal Impact

The literature on the use of journal impact measures uniformly concludes that Gross and Gross (1927) were the first to develop this method (see, e.g., Allen, 1929; McNeely and Crosno, 1930; Gross and Woodford, 1931; Henkle, 1938; Brodman, 1944; Garfield, 1955; and Raisig, 1960). Gross and Gross sought to address the rising problems of small colleges at a time when one "of the biggest of these is

[1] The authors wish to thank Jean-Pierre Robitaille and the two anonymous reviewers for their valuable comments and suggestions. They also wish to thank Julie Caruso and Johanna Kratz for their revisions of the manuscript.

the problem of adequate library facility." It is important to note that the first use of journal impact calculation aimed to facilitate the task of journal selection, which is one core aspect in the marketing of the most visible commercial product that has emerged from this work—Thomson Scientific's Journal Citation Report (JCR).

The paper by Gross and Gross (1927) proved to be an inspiration for several US librarians and early information scientists. For instance, Brodman (1944) cites no less than 18 papers published after 1926 that used a method based on the Gross and Gross paper. It is not surprising that such an explosion occurred. During that period, the number of periodicals available to libraries was growing at an exponential rate, and the Great Depression was taking its toll on the budget of university libraries. Importantly, the problem is just as prominent today. This quote from Cunningham (1935) has all the elements of a librarian's pamphlet from the present day:

> The tremendous number of journals being published and the continued increase in the cost of yearly subscriptions have made it increasingly difficult for libraries to maintain adequate subscription lists. At the same time, libraries have been facing a marked decrease in budgets, gifts and other forms of financial support.

What is immediately obvious when one examines the evolution of journal impact calculations that followed the method pioneered by Gross and Gross is the growing complexity and size of the compilations. 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). The Gross and Gross method grew in size and complexity as it was adopted by other researchers. For instance, the Gross and Gross study carried out in 1927 used a single source journal and comprised a compilation of 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 and Crosno (1930) used seven source journals—"three American, one English, two German publications, and one French publication" (p. 82)—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 fields relevant to medicine and tabulating some 26,700 references from about 40 source journals or monographs. In 1956, Brown 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.

Importantly, all of the aforementioned studies produced field-specific listings, and no author saw a need to adapt the method to enable comparisons across the field. There was, in fact, no need for such cross-field comparisons, since the purpose of the technique was to identify relevant journals for different fields. This characteristic produces some adverse effects when measures of journal impact are used to evaluate scientific production across fields; however, these difficulties could not have been foreseen because, at the time, this type of use was not driving the method's development.

In 1936, Hackh 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. However, this idea, with its added layer of complexity, was not taken up until 1960, when it re-appeared 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," which was coined 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 adopted by Garfield and Sher (1963b) for the calculation of the "journal impact factor". According to Garfield and Sher (1963b), their methodological changes to the existing literature involved the inclusion of multiple citations, in contrast to Raisig's (1960) suggestion, as well as self-citations, thus diverging from Westbrook's (1960) recommended method. Thus, Garfield and Sher's JIF was not a creation *ex nihilo*; it was

essentially a massive scale-up of existing techniques, permitted by the construction of the Science Citation Index in 1961 at the Institute for Scientific Information (ISI).

Interestingly, there is another important characteristic that appeared in Raisig (1960) that was eventually adopted in the construction of the impact factor commercialized by ISI in the 1970s. That characteristic was an asymmetry between the items that were considered valid counts for the numerator and for the denominator. Indeed, Raisig (*Ibid*) mentions that "[e]xcluded from the counts of original articles were letters, review articles, reports of patents, book reviews, abstracts, and purely biographical material." Similarly, a few years later (but possibly without prior knowledge of Raisig's approach), Martyn and Gilchrist (1967) also decided to exclude some source items, such as abstracts, obituaries, reviews, and bibliographies, from the counts. 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 document that are deemed citable) predates ISI's impact factor. The fact that Raisig, as well as Martyn and Gilchrist, were cited by Garfield in 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.

Another important attribute of ISI's impact factor is the controversial two-year citation window that was developed by Martyn and Gilchrist (1967), who clearly exposed this characteristic when they wrote that "68,764 citations were made in 1965 of British items published in 1963 and 1964 to a total of 28,949 items" (p. vii). In a later study, Garfield (1972) mentioned his use of 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. (p. 476)

One can surmise that the use of this approach, based on considering only the previous two years of publications, was largely the result of an accidental choice rather than the result of an in-depth analysis of various solutions and the subsequent choice of optimal characteristics. Indeed, Garfield was aware that, overall, the vast majority of citations were older than two years. In 1963, he wrote that "over 50% of the cited references in the 1961 index are more than five years old" (Garfield and 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 evidence suggests that Martyn and Gilchrist are the creators of the impact factor as we know it. When Garfield (1972) adopted their method and furthered the work undertaken by librarians and information scientists over the course of the previous 40 years, work that had aimed to provide a way of acquiring adequate journals for libraries, his approach actually adopted most of the characteristics that had progressively taken shape in prior work. Like most of his predecessors, whose work had aimed to serve US scientific library users rather than to unambiguously determine the way the scientific system worked at the world level, a large number of Garfield's source journals were no doubt US- and English-centric. Had this method been developed in a different country and had Garfield been, for instance, German, he would have started by using a vast majority of German-language journals as source items. This would have resulted in journal impact values that were different from those he obtained, as well as in 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 impact factor. Had this been the case, the impact factor values presented in the JCR would surely be substantially different.

Consequently, these measures cannot be considered objective measures of the worth of all journals published internationally. The JCR and its journal impact factor measures are the result of historically contingent events, and it is very important to consider the dire impact of these contingencies: the English-language and US bias, the presence of an asymmetrical numerator and denominator, and a

seemingly accidental two-year citation window are all characteristics that have a deep effect on the way research, journals, and even scientists are evaluated around the world today.

Use and misuse of measures of journal impact, and potential remedies

For the sake of convenience, one can categorise 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) if left in the wrong hands, indicators can easily be misused, but there are relatively simple normalizations that can, in general, make their use safer; 3) these indicators are technically flawed, but can be re-engineered in depth, and their flaws can be 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 recognised 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 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 linked with the levels of aggregation. Glänzel and 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 tend to disappear, but there is at least one aspect that nearly everyone who carefully uses these indicators will say: they need to be modified somewhat to take into account the inter-field variations.

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 the first part of this paper made it apparent that these indicators were developed with a clear intent: to support the work of librarians in the management of their journal collections. Likewise, Thomson Scientific has a clear message about the intended uses of the Journal Citation Report[2]:

> 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.

Aside from these intended uses, the JIF often serves in research evaluation. One of the most blatant abuses of this tool involves giving bonuses or making promotion decisions for researchers based on raw impact factor values. For example, Fuyono and 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 their publications. The authors also provide the example of the

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

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 a well-known fact 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 these fields), even if they manage to publish in the best journals in their respective fields. A simple normalization by scientific domain would create a more level playing field and certainly a fairer and more informed reward system. 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[3] (Garfield, 1979, as cited in Schubert and Braun, 1996, p. 312).

There are a very large number of papers with suggestions on how to normalize for differences across fields[4]. 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" (PAGE NUMBER). Even then, some bibliometricians would still advise against this type of usage, given the numerous flaws of the currently dominant JIF.

The **third type** of objection is primarily technically based and is in very large part aimed at the specific incarnation of the JIF commercialized by Thomson Scientific. Unlike the first type of objection, it traditionally came from researchers within the field of bibliometrics but, as can be seen in the editorials of biomedical and clinical research journals, a large number of researchers have something, mostly negative, to say about the limits of this indicator.

An item that is repeatedly criticized is the two-year citation window used in the JCR. The calculation of the impact factor based on a citation window of only two years is far too short in many fields. Glänzel and 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.

Among the technical limits often cited, the asymmetry between the numerator and denominator and journal self-citations are among the most commonly mentioned. This asymmetry induces some strong distortions, more particularly for highly cited journals (Moed and 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 cited than regular articles. This also presents an open invitation to distort the JIF scores by simply increasing the number of reviews. Another way editors can manipulate the JIF is by inducing authors to cite the journal in which they publish, and 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):

---

[3] Following on this idea, Pudovkin and Garfield (2004) suggested using a rank-normalized impact factor using percentiles.

[4] See e.g. Fassoulaki et al. (2002), Huth (2001), Ramirez, Garcia and Del Rio (2000), Schwartz and Lopez Hellin (1996), Sen and Shailendra (1992), Sombatsompop et al. (2005) and van Leeuwen and Moed (2001) to name only a few.

> 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 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)

This last form 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 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 corrected measures for their own research and contractual undertakings, but they would not be authorized to commercialize these indicators and compete against the JCR.

Conclusion

The JCR, which is the most well known descendant of the work undertaken by Gross and Gross in the 1920s, continues to cater to the need of university librarians who have to carefully select the most relevant journals for their clientele within limited budgets. Since the original work of Gross and Gross, measures of journal impact have grown in complexity and in size. There is no doubt that even though the presence of many of the characteristics of the JCR (such as the presence of an asymmetrical numerator and denominator, the use of a two-year citation window and the prevalence of English-speaking journals) are justified with the help of rationale arguments, they are at least in large part the result of historical contingencies. For instance, had the *Institute for Scientific Information* emerged as the "Institut für Forschungsinformation", the JCR would undoubtedly have evolved in a substantially different form.

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 use intended by Thomson Scientific, such as selecting journals for a library, the weaknesses are certainly acceptable. 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, and the deficiencies of the impact factor become increasingly worrisome, for the laws of large numbers decreasingly come into play to compensate for the shortcomings in the way JCR metrics are computed. While some improvements (e.g., field normalization), can be made "in-house" without a huge 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 anytime soon.

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.

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.