

Pros and Cons of the Impact Factor in a Rapidly Changing Digital World

Michael McAleer

Department of Finance Asia University, Taiwan and
Discipline of Business Analytics
University of Sydney Business School, Australia and
Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam, The Netherlands and
Department of Economic Analysis and ICAE
Complutense University of Madrid, Spain and
Institute of Advanced Studies
Yokohama National University, Japan

Judit Oláh

Faculty of Economics and Business
Institute of Applied Informatics and Logistics
University of Debrecen, Hungary

József Popp

Faculty of Economics and Business
Institute of Sectoral Economics and Methodology
University of Debrecen, Hungary

Abstract

The purpose of the paper is to present arguments for and against the use of the Impact Factor (IF) in a rapidly changing digital world. The paper discusses the calculation of IF, as well as the pros and cons of IF. Editorial policies that affect IF are examined, and the merits of open access online publishing are presented. Scientific quality and the IF dilemma are analysed, and alternative measures of impact and quality are evaluated. The San Francisco declaration on research assessment is also discussed.

Keywords Impact Factor, Quality of research, Pros and Cons, Implications, Digital world, Editorial policies, Open access online publishing, SCIE, SSCI.

JEL Classification O34, O31, D02.

Working Paper nº 1806
January, 2018



UNIVERSIDAD
COMPLUTENSE
MADRID

ISSN: 2341-2356

WEB DE LA COLECCIÓN: <http://www.ucm.es/fundamentos-analisis-economico2/documentos-de-trabajo-del-icae> Working papers are in draft form and are distributed for discussion. It may not be reproduced without permission of the author/s.

Pros and Cons of the Impact Factor in a Rapidly Changing Digital World*

Michael McAleer

Department of Finance
Asia University, Taiwan
and

Discipline of Business Analytics
University of Sydney Business School, Australia
and

Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam, The Netherlands
and

Department of Economic Analysis and ICAE
Complutense University of Madrid, Spain
and

Institute of Advanced Studies
Yokohama National University, Japan

Judit Oláh**

Faculty of Economics and Business
Institute of Applied Informatics and Logistics
University of Debrecen, Hungary

József Popp

Faculty of Economics and Business
Institute of Sectoral Economics and Methodology
University of Debrecen, Hungary

February 2018

* For financial support, the first author is grateful to the Australian Research Council and the National Science Council, Ministry of Science and Technology (MOST), Taiwan.

* Corresponding author: olah.judit@econ.unideb.hu

Abstract

The purpose of the paper is to present arguments for and against the use of the Impact Factor (IF) in a rapidly changing digital world. The paper discusses the calculation of IF, as well as the pros and cons of IF. Editorial policies that affect IF are examined, and the merits of open access online publishing are presented. Scientific quality and the IF dilemma are analysed, and alternative measures of impact and quality are evaluated. The San Francisco declaration on research assessment is also discussed.

Keywords: Impact Factor, Quality of research, Pros and Cons, Implications, Digital world, Editorial policies, Open access online publishing, SCIE, SSCI.

JEL: O34, O31, D02.

1. Introduction

Librarians and information scientists have been evaluating journals for almost 90 years. Gross and Gross (1927) conducted a classic study of citation patterns in the 1920s, followed by Brodman (1944), with studies of physiology journals and subsequent reviews following this lead. Garfield (1955) first mentioned the idea of an impact factor in *Science*. The introduction of the experimental Genetics Citation Index in 1961 led to the publication of the *Science Citation Index (SCI)*. In the early 1960s, Sher and Garfield created the journal impact factor to assist in selecting journals for the new SCI (Garfield and Sher, 1963).

In order to do this, they simply re-sorted the author citation index into the journal citation index and, from this exercise, they learned that initially a core group of large and highly cited journals needed to be covered in the new SCI. They sampled the 1969 SCI to create the first published ranking by impact factor. Garfield's (1972) paper in *Science* on "Citation analysis as a tool in journal evaluation" has received most attention from journal editors, and was published before *Journal Citation Reports (JCR)* existed. A quarterly issue of the 1969 SCI was used to identify the most significant journals in science, where the analysis was based on a large sample of the literature. After using journal statistical data to compile the *SCI* for many years, the *Institute for Scientific Information (ISI)* in Philadelphia started to publish *Journal Citation Reports (JCR)* in 1975 as part of the SCI and the Social Sciences Citation Index (SSCI).

However, ISI recognized that smaller but important review and specialty journals might not be selected if they depended solely on total publication or citation counts (Garfield, 2006). A simple method for comparing journals, regardless of size or citation frequency, was needed and the Thomson Reuters Impact Factor (IF) was created. The term "impact factor" has gradually evolved, especially in Europe, to describe both journal and author impact. This ambiguity often causes problems.

It is one thing to use impact factors to compare journals and quite another to use them to compare authors. Journal impact factors generally involve relatively large populations of articles and citations. Indeed, most metrics relating to impact and quality are based on citations data (Chang and McAleer, 2015). Individual authors, on average, produce much

smaller numbers of articles, although some can be phenomenal. The impact factor is used to compare different journals within a certain field. The *ISI Web of Science (WoS)* indexes more than 12,000 science and social science journals.

JCR offers “a systematic, objective means to critically evaluate the world’s leading journals, with quantifiable, statistical information based on citation data” (Thomson Reuters, 2015). However, there are increasing concerns that the impact factor is being used inappropriately and not in ways as originally envisaged (Garfield, 2006; Adler et al., 2009). IF reveals several weaknesses, including the mismatch between citing and cited documents. The scientific community seeks and needs better certification of journal procedures and metrics to improve the quality of published science and social science.

The plan of the remainder of the paper is as follows. Section 2 discusses calculation of the Impact Factor (IF), and the pros and cons of IF are given in Section 3. Editorial policies that affect IF are examined in Section 4. The merits of open access online publishing are presented in Section 5. Scientific quality and the IF dilemma are analysed in Section 6, and alternative measures of impact and quality are evaluated in Section 7. The San Francisco declaration on research assessment is discussed in Section 8. Concluding comments are given in Section 9.

2. Calculation of Impact Factor (IF)

IF is calculated yearly, starting from 1975 for those journals that are indexed in the *JCR*. In any given year, the impact factor of a journal is the average number of citations received per paper published in that journal during the two preceding years. Thus, the impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years (Garfield, 1972). For example, if a journal has an impact factor of 3 in 2013, then its papers published in 2011 and 2012 received 3 citations each, on average, in 2013.

New journals, which are indexed from their first published issue, will receive an IF after two years of indexing. In this case, the citations to the year prior to Volume 1, and the number of articles published in the year prior to Volume 1, are known zero values. Journals that are indexed starting with a volume other than the first volume will not be given an IF until they

have been indexed for three years. IF relates to a specific time period. It is possible to calculate it for any desired period, and the *JCR* also includes a five-year IF. The *JCR* shows rankings of journals by IF, if desired by discipline, such as organic chemistry or psychiatry.

Citation data are obtained from a database produced by *ISI*, which continuously records scientific citations as represented by the reference lists of articles from a large number of the world's scientific journals. The references are rearranged in the database to show how many times each publication has been cited within a certain period, and by whom, and the results are published as the *SCI*. On the basis of the *SCI* and author publication lists, the annual citation rate of papers by a scientific author or research group can be calculated. Similarly, the citation rate of a scientific journal can be calculated as the mean citation rate of all the articles contained in the journal (Garfield, 1972). This means that IF is a measure of the frequency with which the "average article" in a journal has been cited in a particular year or period.

IF could just as easily be based on the previous year's articles alone, which would give even greater weight to rapidly changing fields. A less current IF could take into account longer periods of citations and/or sources, but the measure would then be less current. The *JCR* 'help page' provides instructions for computing five-year impact factors. Nevertheless, when journals are analysed within discipline categories, the rankings based on 1-, 7- or 15-year IF do not differ significantly. Garfield reported on this in *The Scientist* (Garfield, 1998a, b).

When journals were studied across fields, the ranking for physiology journals improved significantly as the number of years increased, but the rankings within the physiology category did not change significantly. Similarly, Hansen and Henrikson (1997) reported "good agreement between the journal impact factor and the overall (cumulative) citation frequency of papers on clinical physiology and nuclear medicine."

IF is useful in clarifying the significance of absolute (or total) citation frequencies. It eliminates some of the bias in such counts, which favor large over small journals, or frequently issued over less frequently issued journals, and of older over newer journals. In the latter case, in particular, such journals have a larger citable body of literature than do smaller or younger journals. All things being equal, the larger is the number of previously published articles, the more often will a journal be cited (Garfield, 1972).

The integrity of data, and transparency about their acquisition, are vital to science. IF data that are gathered and sold by *Thomson Scientific* (formerly the Institute of Scientific Information, or ISI) have a strong influence on the scientific community, affecting decisions on where to publish, whom to promote or hire, the success of grant applications, and even salary bonuses, among others.

3. Pros and Cons of IF

In an ideal world, IF would rely only on complete and correct citations, reinforcing quality control throughout the entire journal publication chain. There is a long history of statistical misuse in science (Cohen, 1938), but citation metrics should not perpetuate this failing. Numerous criticisms have been made of the use of IF. The research community seems to have little understanding of how impact factors are determined, with no audited data to validate their reliability (Rossner et al., 2007).

Other criticism focuses on the effect of the impact factor on the behavior of scholars, editors and other stakeholders (van Wesel, 2015; Moustafa, 2015). The use of IF instead of actual article citation counts to evaluate individuals is a highly controversial issue. Grants and other policy agencies often wish to bypass the work involved in obtaining citation counts for individual articles and authors.

Journal impact can also be useful in comparing expected and actual citation frequencies. Thus, when Thomson Scientific prepares a personal citation report, it provides data on the expected citation impact, not only for a particular journal, but also for a particular year, as IF can change from year to year. Recently published articles may not have had sufficient time to be cited, so it is tempting to use IF as a surrogate evaluation tool. The mere acceptance of the paper for publication by a high impact journal is purportedly an implied indicator of prestige and quality. Typically, when the author's work is examined, the IF of the journals involved are substituted for the actual citation count. Thus, IF is used to estimate the expected count of individual papers, which is seriously problematic considering the known skewness observed for most journals.

It is well known that there is a skewed distribution of citations in most fields, with a few articles cited frequently, and many articles cited rarely, if at all (see Chang et al., 2011). There are other statistical measures to describe the nature of the citation frequency distribution skewness. However, so far no measures other than the mean have been provided to the research community (Rossner et al, 2007). For example, the initial human genome paper in Nature (Lander et al., 2001) has been cited a total of 5,904 times (as of November 20, 2007). In a self-analysis of their 2004 impact factor, Nature noted that 89% of their citations came from only 25% of the papers published, and so the importance of any one publication will be different from, and in most cases less than, the overall number (Editorial, 2005).

IF is based on the number of citations per paper, yet citation counts follow a Bradford distribution (that is, a power law distribution), so that the arithmetic mean is a statistically inappropriate measure (Adler et al., 2008). With a normal distribution (such as would be expected with, for example, adult body mass), the mode, mean and median all have similar values. However, with citations data, these common statistics may differ dramatically because the median calculation would typically be much lower than the mean.

Most articles are not well-cited, but some articles may have unusual cross-disciplinary impacts. The so-called 80/20 phenomenon applies, in that 20% of articles may account for 80% of the citations. The key determinants of impact factor are not the number of authors or articles in the field, but rather the citation density and the age of the literature that is cited. The size of a field, however, will increase the number of “super-cited” papers. Although a few classic methodological papers may exceed a high threshold of citation, many other methodological and review papers do not. Publishing mediocre review papers will not necessarily boost a journal’s impact (Garfield, 2006).

Some examples of super-citation classics include the Lowry method (Lowry et al., 1951), which has been cited 300,000 times, and the Southern Blot technique that has been cited 30,000 times (Southern, 1975). As the roughly 60 papers cited more than 10,000 times are decades old, they do not affect the calculation of the current impact factor. Indeed, of 38 million items cited from 1900-2005, only 0.5% were cited more than 200 times, one-half were not cited at all (which relates to the *PI-BETA* (Papers Ignored - By Even The Authors) metric presented in Chang et al. (2011)), and about one-quarter were not substantive articles but

rather the editorial ephemera mentioned earlier (Garfield, 2006). The appearance of articles on the same subject in the same issue may have an upward effect, as shown in Opthof (1999).

Another aspect is self-citation, in which citations to articles may originate from within a journal, or from other journals. In general, most citations originate from other journals, but the proportion of self-citation varies with discipline and journal. Generally, self-citation rates for most journals remain below 20% (ISI, 2002). It seems to be harmless in many cases, with few editorial citations (Archambaultb and Lariviere, 2009). However, it is potentially problematic when editors choose to manipulate the IF with self-citations within their own journal (Rieseberg and Smith, 2008; Rieseberg et al., 2011).

In addition, the definition of what is considered an “article” is often a source of controversy for journal editors. For example, some editorial material may cite articles (items by the Editor, and Letters to the Editor commenting on previously published articles), thereby creating an opportunity to manipulate IF. In some cases, the Letters section can be divided into correspondence and research letters, the latter being peer-reviewed, and hence citable for the denominator, which can lead to an increase in the denominator and to a fall in IF as Letterstend not to be highly cited.

It has been stated that IF and citation analysis are, in general, affected by field-dependent factors (Bornmann and Daniel, 2008). This may invalidate comparisons, not only across disciplines, but even within different fields of research in a specific discipline (Anauati et al., 2014). The percentage of total citations occurring in the first two years after publication also varies highly among disciplines, from 1-3% in the mathematical and physical sciences, to 5-8% in the biological sciences (van Nierop, 2009). In short, impact factors should not be used to compare journals across disciplines.

The fact that *WoS* represents a sample of the scientific literature is often overlooked, and IF is often treated as if it was based on a census. In reality, *WoS* draws on a sample of the scientific literature, selected following their own criteria (Vanclay, 2012), as amended from time to time (for example, through suspensions for self-citation, although this is not as common as might be expected). Other providers, such as Scopus and Google Scholar, and evaluation agencies (for example, the Excellence for Research in Australia) use different samples of the

scientific literature, so their interpretation of corresponding impact and quality would differ from IF.

WoS policies and decisions to include or suspend a journal also affect IF. For example, World Journal of Gastroenterology was suspended in 2005, so that WoS has no data, but Scopus indicates that the journal had over 6000 citations to articles during 2004-05. Therefore, the suspension of one journal could have deflated IF for other gastroenterology journals by as much as 1%. These sources of variation lead one to question the practice of publishing IF with three decimal points, and to ask why there is no statement regarding variability (Vanclay, 2012). However, the annual JCR is not based on a sample, and includes every citation that appears in the 12,000 plus journals that it covers, so that discussions of sampling errors in relation to JCR are not particularly meaningful. Furthermore, ISI uses three decimal places to reduce the number of journals with the identical impact rank (Garfield, 2006).

WoS and JCR suffer from several systemic errors. A report feature of WoS often arrives at different results from the figures published in JCR because WoS and JCR use *different citation matching protocols*. WoS relies on matching citing articles to cited articles, and requires either a digital object identifier (DOI) or enough information to make a credible match. An error in the author, volume or page numbers may result in a missed citation. WoS attempts to correct for errors if there is a close match. In contrast, all that is required to register a citation in JCR is the name of the journal and the publication year.

With a lower bar of accuracy required to make a match, it is more likely that JCR will pick up citations that are not registered in WoS. Furthermore, WoS and JCR use *different citation windows*. The WoS Citation Report will register citations when they are indexed, and not when they are published. If a December 2014 issue is indexed in January 2015, then the citations will be counted as being made in 2015, not 2014. In comparison, JCR counts citations by publication year. For large journals, this discrepancy is not normally an issue, as a citation gain at the beginning of the cycle is balanced by the omission of citations at the end of the cycle. For smaller journals that may publish less frequently, the addition or omission of a single issue may make a significant difference in the IF.

In contrast, WoS is *dynamic*, while JCR is *static*. In order to calculate journal IF, Thomson Reuters takes an extract of their dataset in March, whether or not it has received and indexed

all journal content from the previous year. In comparison, *WoS* continues to index as issues are received. There are also *differences in indexing*. Not all journal content is indexed in *WoS*. For example, a journal issue containing conference abstracts may not show up in the *WoS* dataset, but citations to these abstracts may count toward calculating a journal IF.

While there may be a delay of several years for some topics, papers that achieve high impact are usually cited within months of publication, and almost certainly within a year or so. This pattern of immediacy has enabled *Thomson Scientific* to identify “hot papers” in its bimonthly publication, *Science Watch*. However, full confirmation of high impact is generally obtained two years later. *The Scientist* waits up to two years to select hot papers for commentary by authors. Most of these papers will eventually become “citation classics”. However, the chronological limitation on the impact calculation eliminates the bias that “super classics” might introduce. Absolute citation frequencies are biased in this way but, on occasion, a hot paper might affect the current IF of a journal.

JCR provides quantitative tools for ranking, evaluating, categorizing, and comparing journals as IF is widely regarded as a quality ranking for journals, and is used extensively by leading journals in advertising. The heuristic methods used by *Thomson Scientific* (formerly Thomson ISI) for categorizing journals are by no means perfect, even though citation analysis informs their decisions. Pudovkin and Garfield (2004) attempted to group journals objectively by relying on the 2-way citational relationships between journals to reduce the subjective influence of journal titles, such as *Journal of Experimental Medicine*, which is one of the top 5 immunology journals (Garfield, 1972).

JCR recently added a new feature that provides the ability to establish more precisely journal categories based on citation relatedness. A general formula based on the citation relatedness between two journals is used to express how close they are in subject matter. However, in addition to helping libraries decide which journals to purchase, IF is also used by authors to decide where to submit their research papers. As a general rule, journals with high IF typically include the most prestigious journals.

IF reported by *JCR* imply that all editorial items in *Science*, *Nature*, *JAMA*, *NEJM*, and so on, can be neatly categorized. Such journals publish large numbers of articles that are not substantive research or review articles. Correspondence, letters, commentaries, perspectives,

news stories, obituaries, editorials, interviews, and tributes are not included in the *JCR* denominator. However, they may be cited, especially in the current year, but that is also why they do not significantly affect impact calculations. Nevertheless, as the numerator includes later citations to these ephemera, some distortion will arise.

Only a small group of journals are affected, if at all. Those that are affected change by 5 or 10% (Pudovkin and Garfield, 2004). According to Thomson Reuters, 98% of the citations in the numerator of the impact factor are to items that are considered as citable, and hence are counted in the denominator. The degree of misrepresentation is small. Many of the discrepancies inherent in IF are eliminated altogether in another Thomson Scientific database called Journal Performance Indicators (Fassoulaki et al., 2002). Unlike *JCR*, the Journal Performance Indicators database links each source item to its own unique citations. Therefore, the impact calculations are more precise as only citations to the substantive items that are in the denominator are included.

Recently, Webometrics has been brought increasingly into play, though there is as yet little evidence that this approach is any better than traditional citation analysis. Web “citations” may occur slightly earlier, but they are not the same as “citations”. Thus, one must distinguish between readership, or downloading, and actual citations in newly published papers. Some limited studies indicate that Web citations are a harbinger of future citations (Lawrence, 2001; Vaughan and Shaw, 2003; Antelman, 2004; Kurtz et al., 2005).

4. Editorial Policies that Affect IF

A journal can adopt different editorial policies to increase IF (Arnold and Fowler, 2011). For example, journals may publish a larger percentage of review articles, which are generally cited more frequently than research reports as the former tends to include many more papers in the extended reference list. Therefore, review articles can raise IF of a journal, and review journals tend to have the highest IF in their respective fields. No calculation of primary research papers only is made by *Thomson Scientific*¹. The numerator restricts the count of

¹ Thomson Scientific was one of the operating divisions of the Thomson Corporation from 2006 to 2008. Following the merger of Thomson with Reuters to form Thomson Reuters in 2008, it became the scientific business unit of the new company.

citations to scientific articles excluding, for example, editorial comment. However, most citations are made by articles (including reviews) to earlier articles (Hernan, 2009).

Journal editors could also cite ghost articles that could usefully increase IF, thereby distorting the performance indicators for real contributors. Given the relatively lax error checking by WoS, it is tempting to include a series of ghost articles in a review of this kind to demonstrate weaknesses of IF (Rieseberg et al., 2011). Some journal editors set their submissions policy as “by invitation only” to invite exclusively senior scientists to publish “citable” papers to increase IF (Moustafa, 2015).

Journals may also attempt to limit the number of “citable items”, that is, the denominator in IF, either by declining to publish articles (such as case reports in medical journals) that are unlikely to be cited, or by altering articles (by not allowing an abstract or bibliography) in the hope that Thomson Scientific will not deem it a “citable item”. As a result of negotiations over whether items are “citable”, IF variations of more than 300% have been observed (PLoS Medicine Editors, 2006). Journals prefer to publish a large proportion of papers, or at least the papers that are expected to be highly cited, early in the calendar year as this will give those papers more time to gather citations. Several methods exist for a journal to cite articles in the same journal that will increase IF (Fassoulaki et al., 2002; Agrawal, 2005).

Beyond editorial policies that may skew IF, journals can take overt steps to game the system. For example, in 2007, the specialist journal *Folia Phoniatica et Logopaedica*, with an impact factor of 0.66, published an editorial that cited all its articles from 2005 to 2006 in a protest against the “absurd scientific situation in some countries” related to use of IF (Schuttea and Svec, 2007). The large number of citations meant that IF for that journal increased to 1.44. As a result of the unedifying increase, the journal was not included in the 2008 and 2009 *JCR*.

Coersive citation is a practice in which an editor forces an author to add spurious self-citations to an article before the journal will agree to publish it in order to inflate IF. A survey published in 2012 indicates that coercive citation has been experienced by one in five researchers working in economics, sociology, psychology, and multiple business disciplines, and it is more common in business and in journals with a lower IF (Wilhite and Fong, 2012). However, cases of coercive citation have occasionally been reported for other scientific disciplines (Smith, 1997; Chang et al., 2013).

Even citations to retracted articles may be counted in calculating IF (Liu, 2007). In an example, Woo Suk Hwang's stem cell papers in *Science* from 2004 and 2005, both subsequently retracted, have been cited a total of 419 times (as of November 20, 2007). The denominator of IF, however, contains only those articles designated by Thomson Scientific as primary research articles or review articles, but Nature "News and Views", among others, is not counted (Editorial, 2005). Therefore, IF calculation contains citation values in the numerator for which there is no corresponding value in the denominator.

5. Merits of Open Access Online Publishing

The term "open access" basically refers to free public access to research papers. Academics have argued that since academic research and publishing were publicly funded, the public should have free online access to the papers being published as a result. Publishing is a highly competitive market, no less so for the open access segment. The big publishers have long recognised the popularity of open access, and now offer a range of publications accordingly. However, somebody always has to pay for publication. This means that the new scientific findings become freely accessible, but researchers generally have to include publication costs in their research budget. Gates Foundation is already going one step further and linking future funding to a requirement of publication under the "creative commons" license, allowing material to be used free of charge for the rapid and widespread dissemination of scientific knowledge.

The strength of the relationship between journal IF and the citation rates of papers has been steadily decreasing since articles began to be available digitally (Lozano et al., 2012). The aggressive expansion of large commercial publishers has increasingly consolidated the control of scientific communication in the hands of 'for-profit' corporations. Such publishers presented a challenge to the open access movement and online publishing, the development of a model of a not for-profit journals run by and for scientists. However, the last decade have revolutionized the landscape of scientific publishing and communication.

For the Open Access movement, the last 15 years have been a pivotal time for addressing the financial and commercial considerations of academic publishing, moving from grass roots

initiatives to the introduction of government policy changes. Over the last decade, there has been an immense effort to change how accessible all of this new (and old) information is to the world at large.

The Hindawi Publishing Corporation seems to have been the first open access publisher. However, PLOS (BioMed Central launched open access in 2000) played a pivotal role in promoting and supporting the Open Access movement. The launch of PLOS had the additional effect of creating pressure on traditional publishers to consider their business models, demonstrating that open access publishing was not equivalent to vanity publishing, even though it is the author who pays the costs associated with publishing in this model. PLOS also showed that open access publishing could be done in a way that might tempt scientists to submit their best work to somewhere other than the established traditional journals. The involvement of PLOS in the Open Access movement has seen the acceptance of open access publishing (Ganley, 2013).

The Fair Access to Science and Technology Research Act in the US has mandated earlier public release of taxpayer-funded research. In the UK, the Research Councils provide grants to UK Higher Education Institutes to support payment of article processing charges associated with open access publishing. The European Commission has a strategy in place that aims to make the results of projects funded by the EU Research Framework open access via either “green” or “gold” publishing. The Australian Research Council (ARC) implemented a policy requiring deposition of ARC-funded research publications in an open access institutional repository within 12 months of publication.

The future for improved access to research is bright. The Howard Hughes Medical Institute, the Max Planck Society and the Wellcome Trust launched in 2012 the online, open access, peer reviewed journal *eLife*, which publishes articles in biomedicine and life sciences. The journal does not promote IF, but provides qualitative and quantitative indicators regarding the scope of published articles. Moreover, articles are published together with a simplified language summary in *eLife Digests* to make them accessible to a wider audience, including students, researchers from other areas, and the general public, which also attracts scientific dissemination vehicles and major newspapers (Malhotra and Marder, 2015).

However, not all forms of open access publishing are equal. A key purpose of providing access is to enable and facilitate reuse of the content, but the licenses publishers use can vary radically from one journal to another. If a paper is open via deposition in a repository, or as part of a publisher's hybrid access model, it may still, unfortunately, remain closed from a reuse perspective.

6. Scientific Quality and the IF Dilemma

It is not surprising that alternative methods for evaluating research are being sought, such as citation rates and journal IF, which seem to be quantitative and objective indicators directly related to published science.

Experience has shown that, in each specialty or discipline, 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 IF. Many of these leading journals existed long before the IF was devised. It is important to note that IF is a journal metric, and should not be used to assess individual researchers or institutions (Seglen, 1997). As the IF is readily available, it has been tempting to use IF for evaluating individual scientists or research groups because it is widely held to be a valid evaluation criterion (Martin, 1996), and is probably the most widely used indicator apart from a simple count of publications. On the assumption that the journal is representative of its articles, the journal IF of an author's articles can simply be aggregated to obtain an apparently objective and quantitative measure of the author's scientific achievements.

However, IF is not statistically representative of individual journal articles, and correlate poorly with actual citations of individual articles (the citation rate of articles determines journal impact, but not vice-versa). Furthermore, citation impact is primarily a measure of scientific utility rather than of scientific quality, and the selection of references in a paper is subject to strong biases that are unrelated to quality (MacRoberts and MacRoberts, 1989; Seglen, 1992, 1995). For evaluation of scientific quality, there seems to be no alternative to qualified experts reading the publications. In the prescient words of Brenner (1995): "What matters absolutely is the scientific content of a paper, and nothing will substitute for either knowing or reading it".

According to Sally et al. (2014), journal rankings that are constructed solely on the basis of IF are only moderately correlated with those compiled from the results of experts. The use of journal IF in evaluating individuals has inherent dangers. In an ideal world, evaluators would read each and every article, and make personal judgments. The recent International Congress on Peer Review and Biomedical Publication held from 8-10 September 2013 in Chicago demonstrated the difficulties in reconciling such peer judgments. Most individuals do not have the time to read all the relevant articles. Even if they do, their judgment would likely be tempered by observing the comments of those who have cited the work. Despite wide use of peer reviews, little is known about its impact on the quality of reporting of published research. Moreover, it seems that peer reviewers frequently fail to detect important deficiencies and fatal flaws in papers.

7. Alternative Measures of Impact and Quality

In the 1990s, the Norwegian researcher Seglen developed a systematic critique of IF, its validity, and the way in which it is calculated (Moed et al., 1996; Seglen, 1997). This line of research has identified several reasons for not using IF in research assessments of individuals and research groups (Wouters, 2013a). As the values of journal IF depend on the aggregated citation rates of the individual articles, IF cannot be used as a substitute for individual articles in research assessments, especially as a small number of articles may be cited heavily, while a large number of articles are only cited infrequently, and some are not cited at all (see Chang et al., 2011). This skewed distribution is a general phenomenon in citation patterns for all journals. Therefore, if an author has published an article in a high impact journal, this does not mean that the research will also have a high impact.

Furthermore, fields differ strongly in their IF. A field with a rapid turnover of research publications and long reference lists (such as in biomedical research) will tend to have much higher IF for its journals than a field with short reference lists, in which older publications remain relevant for much longer (such as fields in mathematics). An average paper is cited ~6 times in life sciences, 3 times in physics, and <1 times in mathematics. Many groundbreaking older articles are modestly cited due to a smaller scientific community when they were published.

Moreover, publications on significant discoveries often stop accruing citations once their results are incorporated into textbooks. Thus, citations consistently underestimate the importance of influential vintage papers (Maslov and Redner, 2008). Moreover, smaller fields will usually have a smaller number of journals, thereby resulting in fewer possibilities to publish in high impact journals. Whenever journal indicators and metrics take the differences between fields and disciplines into account, the number of citations to articles produced by research groups as a whole tend to show a somewhat stronger correlation with the journal indicators. Nevertheless, the statistical correlation remains modest. Research groups tend to publish across a whole range of journals, with both high and low IF. It will, therefore, usually be much more accurate to analyze the influence of these bodies of work, rather than fall back on the journal indicators, such as IF (Wouters, 2013b).

As a result, it does not make sense to compare IF across research fields. Although it is a well known, comparisons are still made frequently, for example, when publications are compared based on IF in multidisciplinary settings (such as in grant proposal reviews). In addition, the way in which IF is calculated in *WoS* has a number of technical characteristics such that IF can be gamed relatively easily by unscrupulous journal editors. A more generic problem with using IF in research assessment is that not all fields have IF as they are only based on journals in *WoS* that have IF.

Scholarly fields that focus on books, monographs or technical designs are disadvantaged in evaluations in which IF is important (Wouters, 2013b). IF creates a strong disincentive to pursue risky and potentially groundbreaking research as it takes years to create a new approach in a new experimental context, during which no publications might be expected. Such metrics can block innovation because they encourage scientists to work in areas of science that are already highly populated, as it is only in these fields that large numbers of scientists can be expected to cite references to one's work, no matter how outstanding it might be (Bruce, 2013). In response to these problems, five main journal impact indicators have been developed as an improvement upon, or alternative to, IF (see Chang and McAleer(2015), among others).

In 1976 a recursive IF was proposed that gives citations from journals with high impact greater weight than citations from low impact journals (Pinski and Narin, 1976). Such a recursive IF resembles Google's *PageRank algorithm*, although Pinski and Narin (1976) use a

“trade balance” approach, in which journals score highest when they are often cited but rarely cite other journals (Liebowitz and Palmer, 1984; Palacios-Huerta and Volij, 2004; Kodrzycki and Yu, 2006). PageRank gives greater weight to publications that are cited by important papers, and also weights citations more highly from papers with fewer references. As a result of these attributes, PageRank readily identifies a large number of modestly cited articles that contain groundbreaking results. In 2006, Bollen et al. (2006) proposed replacing impact factors with the PageRank algorithm.

The *SCImago Journal Rank (SJR)* indicator follows the same logic as Google’s PageRank algorithm, namely citations from highly cited journals have a greater influence than citations from lowly cited journals. The *SJR* indicator is a measure of scientific influence of scholarly journals that accounts for both the number of citations received by a journal and the importance or prestige of the journals where such citations occur, and has been developed for use in extremely large and heterogeneous journal citation networks. It is a size-independent indicator, its values order journals by their average prestige per article, and can be used for journal comparisons in science evaluation processes. SCImago (based in Madrid) calculates the *SJR*, though not on the basis of the Scopus citation database that is published by Elsevier (Butler, 2008).

Eigenfactor is another PageRank-type measure of journal influence (Bergstrom, 2007), with rankings freely available online, as well as in JCR. A similar logic is applied in two other journal impact factors from the Eigenfactor.org research project, based at the University of Washington, namely *Eigenfactor* and *Article Influence Score (AIS)*. A journal’s Eigenfactor score is measured as its importance to the scientific community. The Eigenfactor was created to help capture the value of publication output versus journal quality (that is, the value of a single publication in a major journal versus many publications in minor journals). The scores are scaled so that the sum of all journal scores is 100.

For example, in 2006, *Nature* had the highest score of 1.992. The *Article Influence Score* purportedly measures the average influence, per article, of the papers published in a journal, and is calculated by dividing the Eigenfactor by the number of articles published in the journal. The mean *AIS* is 1.00, such that an *AIS* greater than 1.00 indicates that the articles in a journal have an above-average influence. It does not mean that all relevant differences between disciplines, such as the amount of work that is needed to publish an article, is

cancelled. However, *Eigenfactor* assigns journals to a single category, making it more difficult to compare across disciplines. *Eigenfactor* is calculated on the basis of *WoS* and uses citations to an article in the previous five years, whereas it is two years for *IF* and three years for *SJR*.

Chang et al. (2016) argue that *Eigenfactor* should, in fact, be interpreted as a “Journal Influence Score”, and that the *Article Influence Score* is incorrectly interpreted as having anything to do with the score of an article as *each and every article in a journal has the same AIS*. As a matter of fact, *AIS* is the “per capita Journal Influence Score”, which has no reflection whatsoever on any article’s influence.

The *source normalized impact per paper (SNIP)* indicator improves upon *IF* as it does not make any difference in the numerator and denominator regarding “citeable items”, and because it takes field differences in citation density into account. The indicators have been calculated by Leiden University’s Centre for Science and Technology Studies (CWTS), based on the Scopus bibliographic database that is produced by Elsevier. Indicators are available for over 20,000 journals indexed in the Scopus database. *SNIP* measures the average citation impact of the publications of a journal.

Unlike the journal *IF*, *SNIP* corrects for differences in citation practices between scientific fields and disciplines, thereby allowing for more accurate between-field comparisons of citation impact (CWTS, 2015). *SNIP* is computed on the basis of Scopus by CWTS (Waltman et al., 2013a, b). This indicator also weights citations, not on the basis of the number of citations to the citing journal, but on the basis of the number of references in the citing article. Basically, the citing paper is seen as giving one vote which is distributed over all cited papers. As a result, a citation from a paper with 10 references adds 1/10th to the citation frequency, whereas a citation from a paper with 100 references adds only 1/100th. The effect is that *SNIP* balances out differences across fields and disciplines in citation density.

It is worth mentioning *article-level metrics*, which measure impact at an article level rather than journal level, and may include article views, downloads, or mentions in social media. As early as 2004, the *British Medical Journal* (BMJ) published the number of views for its articles, which was found to be somewhat correlated to citations (Perneger, 2004). In 2008 the *Journal of Medical Internet Research* began publishing views and tweets. These

“tweetations” proved to be a good indicator of highly cited articles, leading the author to propose a “Twimpace factor”, which is the number of Tweets it receives in the, admittedly arbitrary, first seven days of publication, as well as a Twindex, which is the rank percentile of an article’s Twimpace factor (Eysenbach, 2011). Starting in March 2009, the Public Library of Science (PloS) also introduced article-level metrics for all articles (Thelwall et al., 2013).

8. San Francisco Declaration on Research Assessment

It is important that IF be improved, because it is influential in shaping science and publication patterns (Knothe, 2006; Larivière and Gingras, 2010). Several alternative metrics (for example, Eigenfactor, Article Influence Score, hindex: see Chang and McAleer (2015) for a list of citations metrics available for Thomson Reuters), and providers (for example, Scopus and SCImago), are forcing change, and threatening the dominance of IF provided by Thomson Reuters. However, there remains a need for many of the “gate-keeping” services that Thomson Reuters provides in assessing timeliness of publication and the rigour of the review process. This creates the opportunity for Thomson Reuters (or new providers) to reposition such services in a way that is more constructive and supportive of science in evaluating the impact and quality of published papers.

IF had its origins in the desire to inform library subscription decisions (Garfield, 2006), but it has gradually evolved into a status symbol for journals which, at its best, can be used to attract good manuscripts and, at its worst, can be unscrupulously and widely manipulated. IF often serves as a proxy for journal quality, but it is increasingly used more dubiously as a proxy for article quality (Postma, 2007). Despite these failings, in the absence of a clearly superior metric that is based on citations, there remains a general perception that IF is useful and a reasonably good indicator of journal quality.

The value-added that is offered by editors of Thomson Reuters derives from efficient matching of papers with reviewers (Laband, 1990). However, this neglects the editorial role of checking for duplication, “salami” (Abraham 2000), plagiarism, and outright fraud. It is rarely made clear whether this checking is expected of reviewers, and /or completed by the editorial office. Science would be well served by an independent system to certify that editorial processes were prompt, efficient and thorough.

The weakest link in science communication is the certification that establishes that a research paper is a valid scientific contribution. There are several aspects involved, but few of these are an integral part of the review process (Weller, 2001; Hames, 2007). Many of the responsibilities are passed on to voluntary referees, who often lack the time and inclination to check rigorously for fraud and duplicate or “salami” publications (Dost, 2008). Indeed, Bornmann et al. (2008) observe that guidelines for referees rarely mention such aspects. Wager et al. (2009) noted that many science editors seem to be unconcerned about publication ethics, fraud, and unprofessional misconduct.

Some editors seek to push ethical responsibilities back on to the author (for example, Abraham, 2000; Tobin, 2002; Roberts, 2009), despite the prevalence of duplicate and fraudulent publications, indicating that self-regulation by authors is insufficient (Gwilym et al., 2004; Johnson, 2006; Berquist, 2008). There is a potential role for Google Scholar in helping to reduce fraud and plagiarism in science. Google Scholar already routinely displays “*n* versions of this article” in search results, and it could usefully display “other articles with similar text” and “other articles with similar images”. Such an addition would be very useful for researchers when compiling reviews and meta-analyses. Clearly, quality science requires a more proactive role from editorial offices, and the pursuit of this role is most certainly not reflected in any aspect of IF.

IF could be retained in a similar form, but amended to deal with its limitations. Specifically, IF should: (1) rely on citations *from* articles and reviews, *to* articles and reviews; (2) re-examine the timeframe; and (3) abandon the 2-year window in favour of an alternative that reflects the varying patterns of citations accrual in different disciplines. Furthermore, the scientific community could rely on a community-based rating of journals, in much the same as *PLoS One* does for individual articles, and as other on-line service providers offer to clients (Jeacle and Carter, 2011).

Saunders and Savulescu (2008) suggested independent monitoring and validation of research. There have been several calls (Errami and Garner, 2008; Butakov and Scherbinin, 2009; Habibzadeh and Winker, 2009, among others) for greater investment in, and more systematic efforts directed at, detecting plagiarism, duplication, and other unprofessional lapses in the

editorial review process. Callaham and McCulloch (2011) concluded that the monitoring of reviewer quality is even more crucial to maintain the mission of scientific journals.

Despite these many calls for reform, IF remains essentially unchanged, but supplemented with a 5-year variant, and *Eigenfactor* and *Article Influence Score* (recall the caveats about these two measures discussed previously). Thomson Reuters could show strong leadership with a system that is better aligned with quality considerations in scientific publications, including editorial efficiency and constructiveness of the review process. Moreover, procedures to detect and deal with plagiarism, and intentional or unintentional lapses in professional and ethical standards, would be most welcome.

Comparing citation counts to individual journal articles is more informative than weighting the IF values of the journals. For bibliometricians, citation analysis is the impact measurement of individual scholarly items based on citation counts. Citation impact is just one aspect of an article's quality, which complements its accuracy and originality. As a clear definition of scientific quality does not exist, no all-in-one metric has yet been proposed (Marx and Bornmann, 2013). It is well known that citation-based data correlate well with research performance (quality) asserted by peers.

Comparing citation counts in various disciplines and at different points in time can be highly misleading, unless there is appropriate standardisation or normalisation. Normalisation is possible by using reference sets, which assess the citation impact of comparable publications (Vinkler, 2010). The reference sets contain publications that were published in the same year and subject category. The arithmetic mean of the citations for all publications in a reference set is calculated to specify the expected citation impact (Schubert and Braun, 1986). This enables calculation of the Relative Citation Rate (*RCR*), that is, the observed citation rate of an article divided by the mean expected citation rate. As with IF, the calculation of *RCR* has an inherent disadvantage related to the lack of normalisation of citations for subject category and publication year.

Percentiles, or the percentile rank classes method, is particularly useful for normalisation (Bornmann and Marx, 2013). The percentile of a published article gives an impression of the impact it has achieved in comparison to similar items in the same publication year and subject category. Unlike *RCR*, percentiles are not affected by skewed distributions, so that highly

cited items do not receive excessively high weights. Publications are sorted by citation numbers and are allocated to percentile ranks ranging between 0 and 100. The percentile of a publication is its relative position within the reference set, so that the higher is the rank, the greater is the number of citations for the publication. For example, a value of 90 indicates that the publication belongs to the 10% of most highly cited articles. A value of 50 is the median level, which means an average impact. The publication set for the percentiles method ranges from single articles to publication records of an individual scientist or an institution.

Together with percentiles, it is possible to focus on specific percentile rank classes, and particularly on the assessment of individual scientists, with Ptop 10% or PPtop 10% indicators (Bornmann, 2013). Both indicators count the number of successful publications normalised for publication year and subject category. Ptop 10% is the number and PPtop 10% is the proportion of publications that belong to the top 10% most highly cited articles. Given the advantages of the percentiles and related PPtop 10%, the Leiden Ranking and SCImago Institutions Rankings have already incorporated these metrics in the global rankings of academic and research institutions.

The *JCR* have tremendous importance globally, despite a widespread growing demand for more intelligent use of such metrics. The European Association of Science Editors (EASE) published its own statement on inappropriate use of IF in 2007, and is one of the signatories of the San Francisco Declaration on Research Assessment (DORA, 2012). EASE issued an official statement recommending “that journal impact factors are used only - and cautiously - for measuring and comparing the influence of entire journals, but not for the assessment of single papers, and certainly not for the assessment of researchers or research programmes” (EASE, 2007).

In July 2008, the International Council for Science (abbreviated as ICSU, after its former name, International Council of Scientific Unions) Committee on Freedom and Responsibility in the Conduct of Science (CFRS) issued a “statement on publication practices and indices and the role of peer review in research assessment”, suggesting many possible solutions - for example, considering a limit number of publications per year to be taken into consideration for each scientist, or even penalising scientists for an excessive number of publications per year - for example, more than 20 (ICSU, 2008). This will, of course, vary according to discipline and team research, especially in the medical and bio-medical sciences.

In February 2010, the Deutsche Forschungsgemeinschaft (German Research Foundation) published new guidelines to evaluate only articles and no bibliometrics information on candidates to be evaluated in all decisions concerning “performance-based funding allocations, postdoctoral qualifications, appointments, or reviewing funding proposals, [where] increasing importance has been given to numerical indicators such as the H-index and the impact factor” (DFG, 2010). This decision follows similar decisions of the Research Excellence Framework (REF) in the UK. The following is what the REF2014 guidelines have to say about journal IF: “No sub-panel will make any use of journal impact factors, rankings, lists or the perceived standing of publishers in assessing the quality of research outputs” (REF, 2014).

Cawkell, sometime Director of Research at *ISI*, remarked that the *SCI*, on which the impact factor is based, “would work perfectly if every author meticulously cited only the earlier work related to his theme; if it covered every scientific journal published anywhere in the world; and if it were free from economic constraints” (Editorial, 2009).

Scientists at research institutes, funding agencies and universities have a need to assess the quality and impact of scientific outputs. The question arises as to whether scientific output is measured accurately and evaluated wisely. In order to address this issue, a group of editors and publishers of scholarly journals met during the Annual Meeting of The American Society for Cell Biology (ASCB) in San Francisco, USA, on 16 December 2012. The group developed a set of recommendations, referred to as the *San Francisco Declaration on Research Assessment (DORA)*. DORA focuses on IF, and it is a strong plea to base research assessments of individual researchers, research groups and submitted grant proposals on article-based metrics, combined with peer review, instead of on journal metrics.

DORA has garnered support from thousands of individuals and hundreds of institutions, all of whom have endorsed the document on the DORA website. On 13 May 2013, more than 150 scientists and 75 scientific organizations had signed the declaration. DORA has attracted a multitude of comments and responses, including a statement from Thomson Reuters that reiterates the inappropriateness of IF as a measure of quality of individual articles, and encouraging authors to choose publication venues that are based on factors not limited to IF (Thomson Reuters, 2013). Nonetheless, it is unlikely that alternative and more appropriate

citation metrics will soon gain recognition as research assessment tools outside the community of bibliometricians.

The bibliometric evidence confirms the main thrust of DORA, namely that it is not sensible to use IF or any other journal impact indicator based on citations as a predictor of the potential citations of a particular paper or set of papers. However, this does not mean that journal IF does not make any sense at all. At the level of the journal, the improved IF do provide interesting information about the role, position and perceived quality of a journal, especially if this is combined with qualitative information about an analysis of who is citing the journal and in what context, as well as its editorial policies.

Editors generally take the opportunity to analyse of their roles in the scientific communication process, and journal indicators can play an informative role. Furthermore, it also makes sense in the context of research evaluation to take into account whether a researcher has been able to publish in a high quality scholarly journal.

Outputs other than research articles will grow in importance in assessing research effectiveness in the future, but the peer-reviewed research paper will remain a central research output that informs research assessment. Focus should be placed primarily on practices relating to research articles published in peer-reviewed journals, but can be extended by recognizing additional products, such as datasets, as important research outputs by funding agencies, academic institutions, journals, organizations that supply metrics, and individual researchers. This step is needed to eliminate the use of journal-based metrics, such as IF, in funding, appointment, and promotion considerations. Research assessment should be evaluated on its own merits rather than on the basis of the journal in which the research is published.

There is a need to capitalize on the opportunities provided by online publications relaxing unnecessary limits on the number of words, figures, and references in articles, and exploring new indicators of significance, quality and impact. Many funding agencies, institutions, publishers, and researchers are already encouraging improved practices in research assessment. Such steps are beginning to increase the momentum toward more sophisticated and meaningful approaches to research evaluation that can now be established and adopted by all of the key constituencies involved (Dora, 2012).

For research assessment, the value and impact of all research outputs (as well as datasets and software) have to be considered in addition to research publications. This includes a broad range of impact measures and qualitative indicators of research impact, such as influence on policy and practice. A variety of journal-based metrics (for example, 5-year impact factor, EigenFactor, SCImago, h-index, editorial and publication times, among others) can provide a richer assessment of journal quality and performance. Such assessments should be based on the scientific content of an article rather than publication metrics of the journal in which it may have been published. It is argued that decisions about funding, hiring, tenure, or promotion assessments based on scientific content, rather than publication metrics, should be given priority (DORA, 2012).

9. Conclusion

The Impact Factor (IF) is generally used as the primary measure with which to compare the scientific output of individuals and institutions. As calculated by Thomson Reuters, IF was originally created as a tool to help librarians identify which journals to purchase, not as a measure of the purported intrinsic scientific quality of research. However, IF has a number of well-documented deficiencies as a tool for research assessment of quality. Citation distributions within journals are highly skewed, and the properties of IF are field-specific as it is a composite of multiple, highly diverse article types, including primary research papers and reviews. Moreover, IF can be manipulated by editorial policy.

As a number that is calculated annually for each scientific journal based on the average number of times that articles are cited over a specified period, IF is intended to be used as a measure of journal quality than an evaluation of individual scientists. However, scientists are being ranked by weighting each of their publications according to the IF of the journal in which it appeared. The misuse of the journal IF is highly destructive, inviting a gaming of the metric that can bias journals against publishing important papers in fields such as social sciences and ecology that are much less cited than others (for example, biomedicine). Moreover, it can waste the time of scientists by overloading highly-cited journals with inappropriate submissions from researchers who are desperate to gain an IF for their publications.

Improved journal impact indicators and metrics solve a number of problems that have emerged in the use of IF, but all journal impact indicators are ultimately based on a function of the number of citations to the individual articles in a journal. The correlation is, however, too weak to legitimize the application of some journal indicators instead of assessing the inherent quality of the articles.

IF is supposed to address the weaknesses it suffers. Possible improvements include the adoption of a 'like-with-like' basis (that is, citations to articles, divided by the count of articles only), the adoption of a more appropriate reference interval (the present two-year interval is too short for many disciplines), and the introduction of confidence intervals. Procedures that add value and restrict plagiarism and fraud are needed to maintain quality. The future of quality science communication lies in the hands of editors, in particular, and the professions at large, in general.

The IF has a large, albeit controversial, influence in the way published scientific research is perceived and evaluated. IF is a very useful tool for evaluation of journals, but it must be used carefully. Considerations include the number of reviews or other types of material published in a journal, variations between disciplines, and item-by-item impacts. A better evaluation system would involve reading each article for quality, but a simple metric is dedicated to the difficulties inherent in reconciling peer review judgments.

When it comes time to evaluating faculty, most reviewers and assessors do not have the time, or care to take the time, to read the articles. Even if they did, their judgment would be tempered by observing the comments of those who have cited the work. Fortunately, new full-text capabilities in the web make this more practical to perform.

References

- Abraham, P. (2000), Duplicate and salami publications, *Journal of Postgraduate Medicine*, 46, 67-69.
- Adler, R., J. Ewing and P. Taylor (2008), Joint committee on quantitative assessment of research: Citation statistics. [A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS).] *Australian Mathematical Society Gazette*, 35(3), 166-188.
[http://www.austms.org.au/Gazette/2008/Jul08/Gazette35\(3\)Web.pdf#page=24](http://www.austms.org.au/Gazette/2008/Jul08/Gazette35(3)Web.pdf#page=24)
- Adler, R., J. Ewing and P. Taylor (2009), Citation statistics, *Statistical Science*, 24(1), 1-14.
- Agrawal, A. (2005), Corruption of journal impact factors, *Trends in Ecology and Evolution*, 20(4), 157.
- Antelman, K. (2004), Do open-access articles have a greater research impact?, *College & Research Libraries News*, 65(5), 372-382.
- Archambault E., and V. Lariviere (2009), History of the journal impact factor: Contingencies and consequences, *Scientometrics*, 79, 635-649.
- Anauati, M.V., S. Galiani, and R.H. Gálvez (2014), Quantifying the life cycle of scholarly articles across fields of economic research, p. 23 (12 November 2014), *Social Science Research Network (SSRN)*, SSRN-id2542612.
Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2523078
- Arnold, D.N., and K.K. Fowler (2011), Nefarious numbers, *Notices of the American Mathematical Society*, 58(3), 434-437.
- Bergstrom, C.T. (2007), Eigenfactor: Measuring the value and prestige of scholarly journals, *College & Research Libraries*, 68(5), 314-316.
- Berquist T.H. (2008), Duplicate publishing or journal publication ethics 101, *American Journal of Roentgenology*, 191, 311-312.
- Bollen, J., M.A. Rodriguez, and H. Van de Sompel (2006), Journal status, *Scientometrics*, 69, 669-687.
- Bornmann, L., and H.D. Daniel (2008), What do citation counts measure? A review of studies on citing behavior, *Journal of Documentation*, 64(1), 45-80.
- Bornmann L., I. Nast, and H.D. Daniel (2008), Do editors and referees look for signs of scientific misconduct when reviewing manuscripts? A quantitative content analysis of studies that examined review criteria and reasons for accepting and rejecting manuscripts for publication, *Scientometrics*, 77, 415-432.
- Bornmann, L. (2013), A better alternative to the h index, *Journal of Informetrics*, 7(1), 100.

Bornmann, L., and W. Marx (2013), How good is research really? Measuring the citation impact of publications with percentiles increases correct assessments and fair comparisons, *EMBO Reports*, 14(3), 226-230.

Brodman, E. (1944), Choosing physiology journals, *Bulletin of the Medical Library Association*, 32 (4), 479-483.

Butakov, S., and V. Scherbinin (2009), The toolbox for local and global plagiarism detection, *Computers & Education*, 52, 781-788.

Bruce, A. (2013), Impact factor distortions, *Science*, 340 (6134), 787.

Butler, D. (2008), Free journal-ranking tool enters citation market, *Nature*, 451, 6.

Callaham, M., and C. McCulloch (2011), Longitudinal trends in the performance of scientific peer reviewers, *Annals of Emergency Medicine*, 57, 141-148.

Chang, C.-L., E. Maasoumi and M. McAleer (2016), Robust ranking of journal quality: An application to economics, *Econometric Reviews*, 35(1), 50-97.

Chang, C.-L., and M. McAleer (2015), Bibliometric rankings of journals based on the Thomson Reuters citations database, *Journal of Reviews on Global Economics*, 4, 120-125.

Chang, C.-L., M. McAleer and L. Oxley (2011), Great expectatrics: Great papers, great journals, great econometrics, *Econometric Reviews*, 30(6), 583-619.

Chang, C.-L., M. McAleer and L. Oxley (2013), Coercive journal self citations, impact factor, journal influence and article influence, *Mathematics and Computers in Simulation*, 93, 190-197.

Cohen, J.B. (1938), The misuse of statistics, *Journal of the American Statistical Association*, 33(204), 657-674.

CWTS (2015), CWTS Journal Indicators, Centre for Science and Technology Studies, Leiden University, The Netherlands. <http://www.journalindicators.com/methodology>

DFG (2010), Deutsche Forschungsgemeinschaft (DFG), “Quality not quantity” - DFG Adopts Rules to Counter the Flood of Publications in Research Press Release No. 7/23, February 2010. http://dfg.de/en/service/press/press_releases/2010/pressemitteilung_nr_07/index.html

DORA (2013), San Francisco Declaration on Research Assesment (DORA).

Available at <http://www.embo.org/news/research-news/research-news-2013/san-francisco-declaration-on-research-assessment>

[Accessed 17 June 2015].

Dost, F.N. (2008), Peer review at a crossroads - A case study, *Environmental Science and Pollution Research*, 15(6), 443-447.

EASE (2007), EASE statement on inappropriate use of impact factors, *European Science Editing*, 33(4), 99-100.

Editorial (2005), Not-so-deep impact. Research assessment rests too heavily on the inflated status of the impact factor, *Nature*, 435(7045), 1003–1004.

Editorial (2009), Editorial, *Journal of Biological Physics and Chemistry*, 9(4), 139-140.

Errami, M., and H. Garner (2008), A tale of two citations, *Nature*, 451, 397-399.

Eysenbach, G. (2011), Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact, *Journal of Medical Internet Research*, 13(4): e123.

Fassoulaki, A., K. Papilas, A. Paraskeva, and K. Patris (2002), Impact factor bias and proposed adjustments for its determination, *Acta Anaesthesiologica Scandinavica*, 46(7), 902-905.

Ganley, E. (2013), A lot can happen in a decade, *PLoS Biol*, 11(10), e1001689.
<https://doi.org/10.1371/journal.pbio.1001689>

Garfield, E. (1955), Citation indexes to science: A new dimension in documentation through association of ideas, *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, *Journal of the American Medical Association*, 295(1), 90-93.

Garfield, E. (1998a), Long-term vs. short-term journal impact: Does it matter?, *The Scientist*, 12(3), 10-12.

Garfield, E. (1998b), Long-term vs. short-term journal impact (part II), *The Scientist* 12(14), 12-13.

Garfield, E., and I.H. Sher (1963), New factors in the evaluation of scientific literature through citation indexing, *American Documentation*, 14(3), 195-201.

Gross, P.L.K., and E.M. Gross (1927), College libraries and chemical education, *Science*, New Series, 66(1713), 385-389.

Gwilym, S.E., M.C. Swan, and H. Giele (2004), One in 13 ‘original’ articles in the *Journal of Bone and Joint Surgery* are duplicate or fragmented publications, *Journal of Bone and Joint Surgery*, 86-B(5), 743-745.

Habibzadeh, F., and M.A. Winker (2009), Duplicate publication and plagiarism: Causes and cures, *Notfall Rettungsmed*, 12, 415-418.

Hames, I. (2007), *Peer Review and Manuscript Management in Scientific Journals: Guidelines for Good Practice*, Wiley.

Hansen, H.B., and J.H. Henriksen (1997), How well does journal “impact” work in the assessment of papers on clinical physiology and nuclear medicine?, *Clinical Physiology*, 17(4), 409-418.

Hernan, M.A. (2009), Impact factor: A call to reason, *Epidemiology*, 20, 317-318.

Hopewell, S., G. Collins, Ly-Mee Yu, J. Cook, I. Boutron, L. Shamseer, M. Shanyinde, R. Wharton, and D.G. Altman (2014), The reporting of randomized trials: Changes after peer review (CAPRI), *BMJ*, 349, g4145.

ICSU (2008), Publication practices and indices and the role of peer review in research assessment (July 2008), International Council for Science.

<http://www.icsu.org/publications/cfrs-statements/publication-practices-peer-review/statement-publication-practices-and-indices-and-the-role-of-peer-review-in-research-assessment-july-2008>

ISI (2002), Journal self-citation in the journal citation reports - Science Edition (2002), The Institute for Scientific Information (ISI), Thomson Reuters.

<http://wokinfo.com/essays/journal-self-citation-jcr/>

Jeacle, I., and C. Carter (2011), In TripAdvisor we trust: Rankings, calculative regimes and abstract systems, *Accounting, Organizations and Society*, 36(4-5), 293-309.

Johnson, C. (2006), Repetitive, duplicate, and redundant publications: A review for authors and readers, *Journal of Manipulative and Physiological Therapeutics*, 29(7), 505-509.

Kodrzycki, Y.K., and P.K. Yu (2006), New approaches to ranking economics journals, *B.E. Journal of Economic Analysis & Policy*, 5(1). doi:10.2202/1538-0645.1520.

Knothe, G. (2006), Comparative citation analysis of duplicate or highly related publications, *Journal of the American Society for Information Science and Technology*, 57(13), 1830-1839.

Kurtz, M.J., G. Eichhorn, A. Accomazzi, C. Grant, M. Demleither, E. Henneken, S.S. Murray, N. Martimbeau, and B. Elwell (2005), The effect of use and access on citations, *Information Processing & Management*, 41(6), 1395-1402.

Laband, D.N. (1990), Is there value-added from the review process in economics? Preliminary evidence from authors, *Quarterly Journal of Economics*, 105, 341-352.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, et al. (2001), Initial sequencing and analysis of the human genome, *Nature*, 409(6822), 860-921.

Larivière, V., and Y. Gingras (2010), The impact factor’s Matthew effect: A natural experiment in bibliometrics, *Journal of the American Society for Information Science and Technology*, 61, 424-427.

Lawrence, S. (2001), Free online availability substantially increases a paper’s impact, *Nature*, 2001, 411, 521.

Liebowitz, I.S.J. and J.P. Palmer (1984), Assessing the relative impacts of economics journals, *Journal of Economic Literature*, 22(1), 77-88.

Liu, S.V. (2007), Hwang's retracted publication still contributes to Science's impact factor, *Sci. Ethics*, 2, 44-45.

Lozano, G.A., V. Larivière, and Y. Gingras (2012), The weakening relationship between the impact factor and papers' citations in the digital age, *Journal of the American Society for Information Science and Technology*, 63(11), 2140-2145.

Lowry, O.H., N.J. Rosebrough, A.L. Farr, and R.J. Randall (1951), Protein measurement with the folin phenol reagent, *Journal of Biological Chemistry*, 193(1), 265-275.

MacRoberts, M.H., and B.R. MacRoberts (1989), Problems of citation analysis: A critical review, *Journal of the Association for Information Science and Technology*, 40(5), 342-349.

Malhotra, V., and E. Marder (2015), Peer review: The pleasure of publishing, *eLife*, 4, e05770.

DOI: <http://dx.doi.org/10.7554/eLife.05770>. <http://elifesciences.org/content/4/e05770>

Martin, B.R. (1996), The use of multiple indicators in the assessment of basic research, *Scientometrics*, 36, 343-362.

Marx, W., and L. Bornmann (2013), Journal impact factor: "The poor man's citation analysis" and alternative approaches, *European Science Editing*, 39(3), 62-63.

Maslow, S., and S. Redner (2008), Promise and pitfalls of extending Google's PageRank algorithm to citation networks, *Journal of Neuroscience*, 28(44), 11103-11105.

Moed, H.F., Th.N. Van Leeuwen, and J. Reedijk (1996), A critical analysis of the journal impact factors of *Angewandte Chemie* and the *Journal of the American Chemical Society*: Inaccuracies in published impact factors based on overall citations only, *Scientometrics*, 37, 105-116.

Moustafa, K. (2015), The disaster of the impact factor, *Science and Engineering Ethics*, 21 (1), 139-142.

van Nierop, E. (2009), Why do statistics journals have low impact factors?, *Statistica Neerlandica*, 63 (1), 52-62.

Opthof, T. (1999), Submission, acceptance rate, rapid review system and impact factor, *Cardiovasc Research*, 41, 1-4.

Palacios-Huerta, I., and O. Volij (2004), The measurement of intellectual influence, *Econometrica*, 72(3), 963-977.

Perneger, T.V. (2004), Relation between online "hit counts" and subsequent citations: Prospective study of research papers in the BMJ, *BMJ*, 329(7465), 546-547.

Pinski, G., and F. Narin (1976), Citation influence for journal aggregates of scientific publications: Theory with application to literature of physics, *Information Processing & Management*, 12(5), 297-312.

PLoS Medicine Editors (6 June 2006), The impact factor game, *PLoS Medicine*, 3(6): e291.
Postma, E. (2007), Inflated impact factors? The true impact of evolutionary papers in non-evolutionary journals, *PLoS ONE*, 2(10): e999.

Pudovkin, A.I., and E. Garfield (2004), Rank-normalized impact factor: A way to compare journal performance across subject categories, Presented at the American Society for Information Science and Technology Annual Meeting, Providence, RI, 17 November 2004. Published in *Proceedings of the 67th Annual Meeting of the American Society for Information Science & Technology*, 41, pp. 507-515.

REF (2014), Research Excellence Framework (REF), REF2014 guidelines. <http://www.ref.ac.uk/about/guidance/>

Rieseberg, L., and H. Smith (2008), Editorial and retrospective, *Molecular Ecology*, 17, 501-513.

Rieseberg, L., T. Vines, and N. Kane (2011), Editorial - 20 years of molecular ecology, *Molecular Ecology*, 20, 1-21.

Roberts, J. (2009), An author's guide to publication ethics: A review of emerging standards in biomedical journals, *Headache*, 49, 578-589.

Rossner, M., H. Van Epps, and E. Hill (2007), Show me the data, *Journal of Cell Biology*, 179(6), 1091-1092.

Saunders, R., and J. Savulescu (2008), Research ethics and lessons from Hwanggate: What can we learn from the Korean cloning fraud?, *J Med Ethics*, 34, 214-221.

Schuttea, H.K., and J.G. Svec (2007), Reaction of *Folia Phoniatica et Logopaedica* on the current trend of impact factor measures, *Folia Phoniatica et Logopaedica*, 59(6), 281-285.

Seglen, P.O. (1992), Journal impact: How representative is the journal impact factor?, *Research Evaluation*, 2, 143-149.

Seglen, P.O. (1995), Siteringer og tidsskrift-impakt som kvalitetsmål for forskning, *Klinisk Kemi i Norden*, 7, 59-63.

Seglen, P.O. (1997), Why the impact factor of journals should not be used for evaluating research, *BMJ*, 314(7079), 498-502.

Southern, E.M. (1975), Detection of specific sequences among DNA fragments separated by gel-electrophoresis, *J Mol Biol*, 98, 503-517.

Smith, R. (1997), Journal accused of manipulating impact factor, *BMJ*, 314(7079), 461.

Brenner, S. (1995), Loose end, *Current Biology*, 5(5), 568.

Schubert, A., and T. Braun (1986), Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 9(5-6), 281-291.

Thelwall, M., S. Haustein, V. Larivière, and C.R. Sugimoto (2013), Do Altmetrics work? Twitter and ten other social web services, *PLoS ONE*, 8(5), e64841.

Thomson Reuters (2013), Thomson Reuters statement regarding the San Francisco Declaration on Research Assessment.

Available at http://researchanalytics.thomsonreuters.com/statement_re_sfdra/

Thomson Reuters (2015), Journal Citation Reports.

http://wokinfo.com/products_tools/analytical/jcr/

Tobin, M.J. (2002), AJRCCM's policy on duplicate publication, *American Journal of Respiratory and Critical Care Medicine*, 166, 433-437.

Vaughan, L., and D. Shaw (2003), Bibliographic and web citations: What is the difference?, *Journal of the Association for Information Science and technology*, 54, 1313-1322.

Vanclay, J.K. (2012), Impact factor: Outdated artefact or stepping-stone to journal certification?, *Scientometrics*, 92(2), 211-238.

Vinkler P. (2010), *The Evaluation of Research by Scientometric Indicators*, Oxford, UK: Chandos.

van Wesel, M. (2016), Evaluation by citation: Trends in publication behavior, evaluation criteria, and the strive for high impact publications, *Science and Engineering Ethics*, 22(1), 199-225.

Wager E., S. Fiack, C. Graf, A. Robinson, and I. Rowlands (2009), Science journal editors' views on publication ethics: Results of an international survey, *Journal of Medical Ethics*, 35, 348-353.

Waltman, L., and N.J. Van Eck (2013), Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison, *Scientometrics*, 96(3), 699-716.

Waltman, L., N.J. Van Eck, T.N. Van Leeuwen, and M.S Visser (2013), Some modifications to the SNIP journal impact indicator, *Journal of Informetrics*, 7(2), 272-285.

Weller A.C. (2001), Editorial peer review: Its strengths and weaknesses, *American Society for Information Science and Technology Monograph Series*, Information Today, ISBN 9781573871006, pp. 342.

Wilhite, A.W. and Fong, E.A. (2012), Coercive Citation in Academic Publishing, *Science*, 335 (6068), 542-543.

Wouters, P. (2013a), Bibliometrics of individual researchers - The debate in Berlin, *Citation Culture*, page 3, 3 October 2013.

<https://citationculture.wordpress.com/2013/10/03/bibliometrics-of-individual-researchers-the-debate-in-berlin/>

Wouters, P. (2013b), The evidence on the journal impact factor, Citation Culture, page 2, 3 June 2013.

<https://citationculture.wordpress.com/2013/06/03/the-evidence-on-the-journal-impact-factor/>