Scientometrics (2012) 92:485–503
DOI 10.1007/s11192-012-0697-6

The journal impact factor: angel, devil, or scapegoat? A comment on J.K. Vanclay's article 2011

Michel Zitt

Received: 22 February 2012 / Published online: 15 March 2012

Abstract J.K. Vanclay's article is a bold attempt to review recent works on the journal impact factor (JIF) and to call for alternative certifications of journals. The too broad scope did not allow the author to fulfill all his purposes. Attempting after many others to organize the various forms of criticism, with targets often broader than the JIF, we shall try to comment on a few points. This will hopefully enable us to infer in which cases the JIF is an angel, a devil, or a scapegoat. We shall also expand on a crucial question that Vanclay could not really develop in the reduced article format: the field-normalization. After a short recall on classical cited-side or ex post normalization and of the powerful influence measures, we will devote some attention to the novel way of citing-side or ex ante normalization, not only for its own interest, but because it directly proceeds from the disassembling of the JIF clockwork.

Keywords Bibliometric measures · Impact factor · Impact factor limitations · Field-normalized impact-factor · Citation behavior · Citation normalization · Citing-side normalization · Source-level normalization

Introduction

J.K. Vanclay's[1] recent article Vanclay (2011) on the journal impact factor (JIF) is a new occasion to debate about the commander's statue in bibliometrics. The statue is still standing in spite of injuries from bad weather, nesting birds, grape shot, and attempts to unscrew it. The monument is paradoxically outdated and timely. Outdated because the

[1] JKV in the followings.

M. Zitt (✉)
INRA Lereco (SAE2) U1134, Rue de la Géraudière, BP 71627, 44316 Nantes Cedex 03, France
e-mail: michel.zitt@nantes.inra.fr

M. Zitt
Observatoire des Sciences et des Techniques (OST), 75015 Paris, France

particular question it raises, namely the journal evaluation, can now be addressed in better ways, either by patches on its weak points or by more radical solutions; timely for a good reason, in that it keeps inspiring those novelties, and also for a bad one, since it continues to be used again and again in its original form, mostly for deviating purposes. Although uncomfortably installed between a review on the JIF and a contribution on journal certification, JKV's work states or restates a series of good questions and is appropriate for fueling the debate.

JKV's article boldly attacks the topic, with quite a broad scope: summarizing the debate on JIF; keeping track of some novelties in context; advocating alternatives for journal certification, all this in a single article's format. If we play the robot-bibliometrician-counting pages and items, we reckon about 220 references for 17 pages full-text. Grasping both the purpose of a review and of a research article is a challenge: many items are just listed with their theme, along with a general classification partly inspired by bibliometric mapping (Table 2). This table commands both admiration on extensiveness and frustration on superficiality. Although we ourselves did plead for the robustness of bibliometric mapping by comparing word and citation-based techniques, we doubt very much that a review can ask so much of co-citation maps. Besides, the announced focus on recent research is justified for new frontiers but suffers from a lack of hindsight on continuous thematic lines.[2] Unfortunately, despite the length of the bibliography, some new or renewed research lines are hardly commented: for example the new prospects on influence measures, normalization, or on formal properties of indicators. The core of the contribution, from the research point of view, might be the promising alternatives for journal certification, in reaction to the impact factor limitations, but the argument is eaten up by the details of the tentative review. Some nice pathways are outlined especially along Web 2.0, so one could expect those perspectives to be critically questioned with the same rigorous requirements as the JIF. How does one certify certifications?

A strange thing about the JIF is that criticism from scientists and other users often attacks scientometrics on its weak point, the JIF, while most bibliometricians, for three decades at least, have warned the scientific community, time and again, about its (mis)uses for evaluation purposes. It remains that probably all of them pay tribute to JIF as a milestone in the field: it was an outstanding innovation at its time (Garfield and Sher 1963; Garfield 1972); very soon it inspired powerful variants of citation analysis, the best example being Pinski and Narin (1976) influence weight; lastly, for the best and the worst, JIF has established the durable matching of scientometrics and the science system.

The JIF deserves a fair trial: this means not confusing its ambitions and its misuses, and not confusing JIF as such, with its claims and technicalities, and JIF as a symbol of (pick) citation analysis, scientometrics as a whole, evaluation system, organization and market of scientific communication. The first section is devoted to JIF's trial and variants or alternatives for journal evaluation, the second to a particular issue: the point of making impacts and citations comparable across fields through adequate normalization. We shall try, inasmuch as possible, to quote other references than JKV's except for some classical works, among them some synthetic presentations such as Glanzel and Moed (2002).

[2] For example in the category "distribution", one might infer that the non-normality is a recent finding, whereas it is the basis of one of the most cited papers about the JIF (Seglen) put in another category. Fortunately, some older references resisted the filtering.

JIF's trial

Let us start with observations that are common knowledge.

1. Garfield and Sher (op.cit.) with forerunning papers back to the 50's (Garfield 1955) introduced a measure of journal notoriety based on average citations per paper, and the power of the company he founded, the Institute for Scientific Information (ISI), allowed this measure to be released and diffused among libraries and labs all over the world, through the journal citation report (JCR). In most progressive areas of science, institutions and laboratories learned to use this new indicator firstly for driving their publication strategy.
2. Among alternatives to JCR, the finer grain Science Citation Index was much more difficult to handle in the early period. It was tempting for institutions and policy makers to use the JIF as a proxy of articles visibility, and furthermore as a proxy of scientists' performance. Thus JIF soon became part not only of publication strategies but of scientific evaluation and perhaps recruitment processes, explicit or tacit: it was the first large-scale application of scientometrics to scientific life, illustrating sociologists' publication/career cycle with a scientometric input (Wouters 1997).
3. Path-breaking measures may convey both novel principles and a specific technical form. As a milestone first of citation analysis and then of general scientometrics, the JIF tends to focus general criticism on both.
4. As a symbol of journal-based evaluation, JIF also attracts criticism about the way journals achieve their role in the scientific system: peer review validation, diffusion, archival.

So, when investigating the case of JIF, the first thing to examine is the target of attacks: is it the JIF as such? Or the context of the JIF (citation analysis, underlying database, role of journals of the communication system)? In the latter case, should the JIF endorse the sins of the fathers, or is it rather an expiatory victim of misuses? Then we can decide whether the JIF is an angel, a devil, or a scapegoat.

The JIF context

Citation analysis

The biases and dangers of citation analysis have given rise to an overwhelmingly abundant literature in sociology and scientometrics and occasionally in economics, management sciences, life sciences and physical sciences. The viability of autonomous theories of citation is questionable. They are mostly embedded in powerful theoretical programs in sociology of science and research (Merton 1942; Hagström 1965; Bourdieu 1975; Callon and Latour 1981). Citation studies also benefit from empirical studies, especially typologies, helping for example to distinguish perfunctory references from cognitive references. However, the abundance of the theoretical material from competing schools of thought, of empirical material with disparate frameworks and various definitions, does not necessarily shed light on the question of "what a citation is". We shall not even browse those well-known points carefully described in many articles (Luukkonen 1997) and reviews (Cronin 1984). Data producers' excessive claims do not help: Adler et al. (2008) quote quite a bold statement by Thomson Reuters which goes beyond pragmatism: "*By recognizing that the value of information is determined by those who use it, what better way to measure the*

quality of the work than by measuring the impact it makes on the community at large. The widest possible population within the scholarly community (i.e., anyone who uses or cites the source material) determines the influence or impact of the idea and its originator on our body of knowledge.[3]" This kind of statement, which imprudently mixes quality and scholarly circulation, is surely not appropriate for defending citation analysis.

On the other hand, the formalization of citation and referencing helps to understand other properties within general models of informetrics. In the mountain of arguments about citation analysis, let us pick just two recurrent questions:

The scale issues:

- In the foreword of *Little Science, Big Science* de Solla Price (1963) insisted on the aggregate view underlying the quantitative analysis of science, in this respect comparable to thermodynamics. Literature reports heavy biases when using authors' citation counts alone in evaluation. A fair correlation between peer review and bibliometric performance, factors not that independent, means that a disagreement exists for a large proportion of individuals. Even if part of the unexplained variance should be put down to inefficiency of peer review rather than of citation analysis, the risk of using citation score only may be unbearable in personal evaluation. Yet, the same skew laws that establish hierarchy in science produce interferences between levels of observation, especially in extreme cases of concentration. The tails of static citation distributions, when modeled by scale-free Zipf-Pareto model, seem to remain on the safe side, with a finite variance, although the Pareto exponent depends on choices on the tail cut-off, on the citation time-window, and on the discipline. Another sign of a tamed distribution is that the aggregate level (journal) departs from Pareto. Though these distributions are usually secured from wild random, severe effects may happen: a star in a lab may collect as many citations are the rest of the staff. This suggests various questions: are straight citation counts misleading and should they be rescaled, compensating for the Matthew effect? Could the dependence on the presence of a star make predictions quite uncertain at an aggregate level (the dynamics of a lab, of a research front…), while betting on the future productivity of the star might be safer[4]? This is perhaps an example of the phenomenon of blurred levels shown by Callon and Latour in a famous paper (*op.cit.*). The issues of scale are pervasive in infometrics, in direct relation to the Lotka-Bradford-Zipf trilogy.
- *Citation and impact factor, novelty killers?* Many scientists react aggressively to possible counter-productive effects of citation-based evaluation. There is a wide choice of such reactions, some found in JKV's bibliography, and debates have been largely echoed in articles or forums in Nature, Science, PNAS. The open book of complaints p. 9–10 could be indefinitely expanded. We shall quote a comment from a neurobiologist,[5] which is

[3] http://scientific.thomson.com/free/essays/citationindexing/history/.

[4] Individual-level productivity modeling is now partly directed towards the predictive capability of the h-index (Hirsch 2007), a matter of controversy (Hönekopp and Khan 2011). In certain areas (biology), the presence of star scientists is not only held as a predictor of future scientific success but also of industrial success and radical changes (in biology: Zucker and Darby 1996).

[5] Four conjectures on impact factors (Jacques Ninio)Each statistical indicator entails biases which need to be identified and corrected. However, in the evaluation of scientific research, popularity tests are used as substitutes for quality tests, a practice which penalizes our most original productions.Conjecture 1: The impact factor of a journal is directly correlated with the incompetence of those who cite it [this argument against multidisciplinary journals]Conjecture 2: Impact factors are directly correlated with lack of originality. Conjecture 3: The impact factor of a journal is inversely correlated with the longevity of the articles it publishes. Conjecture 4: The impact factor of a journal is directly correlated with its rate of fraudulent articles. Source: www.lps.ens.fr/∼ninio/impact-factor.htm.

emblematic of harsh judgments often heard among researchers, and the more solemn IMU's[6] report by Adler et al. which voices the suspicion of mathematicians against citation analysis (*op. cit.*). Beyond over-reaction there is a real worry about the reductive power of citations, especially in the impact factor form, and its potential nuisance for discouraging creativity, exploration of risky research ideas and long-term investment. This is probably the most serious risk about citation analysis and bibliometric indicators, which is not without relation to the predictive capability of "performance" indicators above.

It should be recalled, however, that many traps of evaluation-oriented citation analysis are spared or mitigated in the second family of applications: the mapping-oriented citation analysis.

To sum up, JIF being the archetype of citation analysis, it must endorse its limitations, although sharing the burden with many other applications. It cannot be charged for the faults of parallel applications; conversely, citation analysis should not be charged for weaknesses specific to the JIF. Awkward statements confusing visibility through citation and quality of research should be avoided. The JIF should plead guilty for settings such as the short timeframe that make it particularly sensitive to citation traps.

Database coverage and limitation

The JIF endorses the coverage of the underlying database, "The Web of Science". This classical issue, both theoretical and operational, again emphasized by the JKV (p. 6), concerns all citation or bibliometric applications based on Thomson Reuters data. The author could still have been still more persuasive on this critical point by comparing Google Scholar, Scopus, and WoS on broad subjects like computer science. A most detrimental issue to scientometric analysis is the variability of results: scientists' individual impact or h-index by database. JKV provides an example for journals Table 3, more cruel is the comparison by Bar-Ilan (2008) at the individual level. The discrepancy is not a disaster if it reflects different definitions of the set of citing sources which are explicit, consistent and realistic. The severe gaps of the WoS for non-article sources have been recurrently pointed out, even putting apart the critical case of social science and humanities, source of an abundant literature especially since Hicks (2004). Over the years, Thomson Reuters has taken steps to cover conferences, with many cycles now integrated in the WoS. The coverage of books is the next step, with the recent announcement. Since the beginning Scopus made efforts to cover conference proceedings. Comparisons between databases, or between databases and national/international selections of sources for the purpose of evaluation, are also available.

At the same time, the question of a quasi-Bradfordian selection, discarding marginal sources, as far as international benchmarking is concerned, is a real issue for bibliometric exploitation of the databases. The question is intricate: feeling pressured by some users and

[6] In Adler, Ewing and Taylor, Report of the International Mathematical Union (Adler et al. 2008; see also Adler et al. 2009): "While numbers appear to be 'objective', their objectivity can be illusory. The meaning of a citation can be even more subjective than peer review. The sole reliance on citation data provides at best an incomplete and often shallow understanding of research—an understanding that is valid only when reinforced by other judgments. Numbers are not inherently superior to sound judgments."
Source: http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf.

offers of web engines, the producers show a tendency to expand the set of sources. Good news for the extension of coverage, general information retrieval, capture of various dimensions of lab's activity, including grey literature; bad news for the relevance of coverage on a major purpose, monitoring of international science, but it is up to scientometricians to do their own selection. The qualification of the set of citing sources is a key element. Thomson Reuters keeps claiming that the selection of citing sources is vital against web engines policy but the trade-off between coverage and qualification is challenging (see below).

Another issue is the quality and accuracy of data (capture errors, mismatches, and errors propagation) already repeatedly addressed in literature, at least since Moed and Vriens (1989). We shall not expand on the various facets of these problems, which are well documented but need periodic updates, hopefully recording progress in data handling and unification by database producers. Comparative analysis of databases is welcome.

The JIF cannot exonerate from the limits of the underlying database, just like other bibliometric measures. However, the arguments should not be mixed up; a good principle may be applied to unsatisfactory data and vice versa.

Application of JIF to article or author-level evaluation

Perhaps the most commented-on issue is the diversion from journal-level evaluation to article-level and author-level evaluation. This confusion of levels has been recurrently stressed as a mortal sin since Seglen (1992). The skew distribution and the blurring of scales, again, make the JIF, moreover, built as a mean, quite dependent on the visibility of a few individual articles. Going astray from the journal-level is the typical misuse of which JIF, if we guess its terms of reference, is not guilty. Having said, after many others, that the JIF is hopeless for proper paper or individual evaluation, let us now bring in a nuance.

Ghost bibliometrics: Suppose that the *Journal of Errors and Truth* enters the database and, by a mistake that escapes both the producer and the scrutiny of bibliometricians, receives a JIF which is five times its real value, for several consecutive years. This may cause quite a few authors to anticipate rewards from submitting manuscripts. The process triggers competition. The editors, not willing to make their journal too big, play the card of a stringent selection, recruiting a few high-level scientists in the board. A few years after, when the mistake is eventually corrected, the real JIF has rejoined or overpassed its fake level. This caricature example of self-fulfilling process just helps to illustrate that competition for access to journals, spurred by the diffusion of impact measures, is part of the scientific game.

The simplest citation-base strategy of publication is the trade-off between rejection risk and associated delays on the one hand, and prestige and/or citation expectation on the other, even crudely estimated by a mean measure. Passing the barrier of reviewing in prestigious journals is also held as an achievement. In addition, in the eyes of many users, time-series of journal impact values are, for most journals, easier to extrapolate than individual scattered series, especially for recent articles with theoretical limits stated above.

These aspects of simplicity and convenience may explain the survival on impact factor among users. Most scientometricians, in contrast, soon turned towards the "real citations" data at the paper level, the Science Citation Index made available, also, by Garfield. They also defended the sensible position of decoupling two forms of competition: for access to

journals of good visibility, and for visibility within a given journal.[7] The first factor expresses the actor's position in across-journal competition, the second in within-journal competition (on the topic see also van Raan 2001). There may be a discipline effect in the respective perception, among laboratory heads or evaluators, of actors' impact factor and total impact.

Even assuming the principle of citation analysis is sound with many precautions, the application of journal-level indicators, on its own, to individual evaluation, remains an unjustifiable and outdated practice. The nuance is that the capability of investing visible journal is a component, likely to vary among discipline, of publication and career strategies.

The role of journals in scientific communication

Journals are considered central to the life of scientific community, with at least three missions: validation, diffusion, and archival. The requirement that journal publishing should undergo quality control processes, including fraud and plagiarism control, is of utmost importance and meets the expectations of the scientific community (p. 17–18). The ethics of publication is abundantly debated as JKV recalls, and there are indications that, for example, retractions are multiplying, for one part due to misconducts. International committees (COPE) are proposing guidelines; publishers are now concerned by ethics on authoring and reviewing. JKV rightly stresses the availability of plagiarism detection tools that should be part of the reviewers' toolkit, to be provided either by the journal editors or the publishers.[8] Detection of falsification of results is a much more difficult task.

The particular argument we have some difficulty catching is the role that JKV expects from Thomson Reuters in those process, on the ground of the "gatekeeping implicit in the inclusion in WoS". ISI then Thomson Reuters, the latter firstly a newspapers publisher are not primarily publishers of scientific journals, but intermediary actors in the channel of communication; Elsevier with Scopus plays both roles, hopefully separated. Quality control is an editorial responsibility, not the burden of those actors. What a second-level producer should do is to post the journals' label if a certification exists. Of course, scientometric tools mastered by Thomson Reuters or Elsevier may play a role in the toolbox of quality checking, and the more control and information provided by second-level producers, the better. The lack of communication on key elements such as the rejection rate[9] is another issue, but how can one rely on the sincerity of editors' statements on a strictly internal process if some of the same editors, not respecting the ethics rules they demand from authors/reviewers, show dubious practices of JIF manipulation?

The JIF should be questioned on the ground of its own features and not as a scapegoat of all dysfunctions of the journal-based communication system in science.

[7] A classical way to express this is to see an actor's real impact as the product of the "expected impact" (the actor's impact factor, only depending on journals of publication) and the "relative citation rate" (RCR, Schubert and Braun 1986), particular form of journal-level normalized indicator with neutral value 1 (non-unique decomposition). A RCR type indicator, at another level normalization (fields), is the CWTS "crown indicator" in its original form.

[8] At least do discourage copy-paste and blunt forms of plagiarism. Sophisticated forms may trigger a race, in computational linguistics applications, between anti-plagiarism tools and text transformation tools.

[9] Quite difficult to interpret, however, if only because of authors' anticipations.

Specific issues of the JIF

Then we are left with the specific issues of JIF: Is JIF a good measure of a journal's quality or notoriety?

The devil in disguise

More severe issues are:

- The framework of JIF calculation ("synchronous view" using a citing year base rather than cited year base; handling of relations at the aggregate level; absence of "square design", see below) is a sort of singleton in modern bibliometrics. JKV is totally right on this point, and is probably aware that most published indicators for decades by bibliometrics institutions in various countries are based, in contrast with the JIF framework, on 1–1 relation between a cited article and its citers.[10] Their usual presentation is by "cited year" (diachronous[11]), with the advantage of needing less hypotheses on the stability of journals. The journal demography issue is present whatever the framework. In the JIF, the absence of a strict square design forcing, as much as possible, the same selection principles on the citing and cited-side, is prone to real flaws. The absence of homogeneity for the type of document was raised by Moed and Van Leeuwen (1995) and repeatedly studied since.
- The statistical features: underlying paper-level distributions of citations are approximately Pareto, at least in the tail. Means are part of the shape parameter in the theoretical distribution. However, in distributions with fat tail, empirical means are quite dependent on the occurrence of individuals with large deviations in other words; indicators based on empirical means are quite dependent on extreme values. To a certain extent this is observed for journal impact, which also jeopardizes time series. Confidence intervals also need a proper frame remembering that the WoS is absolutely not a statistical sample for example by considering annual journal data as samples from a multiyear trend.

Descriptive statistics offer a variety of alternatives: the log-normal approximation may suggest a geometric mean, to correct for the non-cited; similarly for the harmonic mean; the mode, suggested by JKV, is not really discriminant, often trapped in the low-cited range with a large number of ties. In the competition between cardinal and rank measures, evaluations based on higher quantiles, typically Q90 and Q80, are better candidates than the median. They are easy to interpret; the choice is a matter of trade-off between discriminating power (in this respect the higher the quantile, the better) and robustness (the higher the quantile, the most fluctuating). Journal evaluations derived from the h-index (Braun et al. 2006) have received a lot of attention, with their usual limitations, size-dependence and difficulty of field-normalization. Related composite indicators are appealing (Glänzel 2008).

As are many other measures, the use of single indicators to represent complete distributions is questionable, as recurrently stressed (Adams et al. 2007; Adler et al. 2008, *op.cit*). This also appears in recent works on axiomatic of citation and impact measures

[10] OST Paris made this choice, for example, in 1992, after other producers.

[11] On a systematic treatment of these differences, see for example Ingwersen et al. (2001).

(see e.g., Rousseau 2008; Marchant 2009) which completes or contradicts intuition. This trend is neglected by JKV, although it is true that for the time being few works in this line address the JIF as such. Bouyssou and Marchant (2011) propose a generalized impact factor based on expected utility. Nice properties, such as increasing the indicator (Marchant's "independence" or Waltman' "consistency", Waltman and Van Eck 2009) when adding a publication, are often trivially obtained for size-dependent indicators (total citation, h-index…).

Robustness against mischievous uses:

– No bibliometric (or economic) indicator can pretend to be invulnerable. Especially, manipulation can exploit internal defects, for example the issue of "types of documents" in the JIF.

The devil in details

– The citation window in the canonical JIF (2 years, before the citing year) is obviously insufficient in slow disciplines. Thomson Reuters now releases an additional JIF calculated on a 5 years window. A shortcoming is the enhanced probability of events in journal demography: merges, splits, and changes.
– The illusory precision of the measure (number of digits) is a good point by JKV. Garfield (2006) already explained that it was a pure convenience to avoid ties, but such levels of precision are unrealistic in the context.

More interesting is the first point. In the same text, Garfield minimizes the effect of the short-window for intra-field comparisons not for across-field comparisons. Yet, as a common use of the JCR is the evaluation of particular journals, the statistical argument of a good level of correlation between rankings using various window lengths is questionable: what if users are interested in those numerous journals that fall in the unexplained part of variance?

Similarly, the "by and large" argument about JIF series' general agreement with the perceived quality of journals (Garfield, *ibid*. citing Hoeffel 1998) is hardly sustainable. The widespread diffusion of the JCR creates expectations among thousands of users and scholars willing to drive their publication policy, and they might be misled by the exceptions. No measure is perfect of course, but defects like the unsatisfactory framework or, for a long period, the short time-frame, are noteworthy: in such matter, the devil is in details.

The JIF could plead guilty on those points, but not for claiming exclusivity. Other measures existed, however, the space of competition was not really open for decades due the quasi-monopoly and market power of ISI and then Thomson Reuters. This is no longer the case.

Among many complements or alternatives, apart from non bibliometric approaches (experts' sayings, users' surveys, usage measures, etc.; see for example Bollen et al. 2009):

(a) Citation based

– *Directly based on the JIF*: normalized impact factor by classical ex-post methods (see below);
– *Diverted from the JIF*: new normalized impacts (see below); influence weights, Eigen factor.
– *Other network-based approaches*: centrality and betweenness measures, etc.

- Other citation measures; total citations (size-dependent).
- *Composite measures*: journal-level h-index and family (obviously size-dependent[12]).
- *Enhanced measures*: influence measures, citing-side normalized measures.

(b) Other aspects

- *Publication-based*: volume, type of documents.
- Internationalization rates at several levels (publication profile; citation profile; editorial committee profile; co-publication profile) and for several geopolitical grids (nations, groups of nations, and continents).

(c) Publishers-level *Rejection rates*: scarce studies, seldom publicly available.
JIF has limitations and at least one flaw.

JKV proposals (options p. 18)

Option 1

Several proposals in this option are perfectly sound, the first one especially. bibliometric bureaus/observatories that recalculate citation impact at all levels (nations, institutions, and individuals) starting from paper-level data (WoS), have implemented correct approaches for decades and mostly rely on a one-to-one and "cited year" approach.

Next, the author recommends "*To abandon the 2-year window in favor of an alternative that reflects the varying patterns of citation accrual in different disciplines*". As already mentioned, Thomson Reuters finally added the 5 years window. The citation obsolescence issue ("aging") is a classics in scientometrics and its variation across fields pleads for a disciplinary modulation of the citation window (Vanclay 2009; Vieira and Gomez 2011) insofar as one stays in the realm of traditional "ex post" normalization. The proposal endorses the fundamental problems of the traditional approach, and the argument in favor of a modulation loses a large part of its strength if one turns to "ex ante" (citing-side) normalization. We discuss these points in the next section.

Citing sources should be submitted to some qualification process and the "square design" (same rule of selection on both sides) has the merit of simplicity (see also Zitt 2011). JKV's proposal is, however, inapplicable as such:

- He generalizes a situation which prevails in most life science and physical science discipline but which is far from general. A stringent universal restriction to articles and reviews[13] would jeopardize the indicators in computer science and related areas, where conference proceedings play a crucial role, both as literature items and citation source. Thomson Reuters has for decades faced criticism of users and bibliometricians on this coverage issue, and progressively shaped the CPCI (and now the Books Index) to finally merge it within the WoS. Scopus tried to integrate proceedings since the beginning. In several fields of social science and humanities, the same problem is met for books and sometimes for proceedings too. As far as computer science and

[12] The statistical relation of the h-index to JIF and size is studied in Glänzel and Schubert (2007).

[13] Bibliometric usage tends to include note and letters, depending on the database, as citable documents. Especially for border types, the qualification of a given document may differ between the journal publisher and the databases and among databases.

engineering fields are concerned, ruling out big cycles of computer science conferences, for example, while a vast majority of specialists have been crying out for their integration, would be an incomprehensible set back – especially for indicators at the institutional and individual level. The solution is not straightforward. A few conference cycles can even compete in notoriety with prestigious journals, but on the other side the huge inequality of conferences prestige makes it necessary to edict some qualification rule as citation sources.

– The statement "The fact that WoS is a sample of scientific literature is often overlooked" (p. 15) will probably astonish quite a few bibliometricians, unless "sample" is understood in a most lax and informal use, and not in any statistical sense. Qualifying the WoS as a sample in statistical terms is deeply misleading. In well-covered disciplines at least (natural sciences), the WoS was conceived as, claims to be, and largely is, a selection of the most important journals in Bradford's spirit sensu lato on the cited-side, and a bit differently on the source side. We can put it as: a census rather good[14] in the high tail, pretty good in the center of the distribution, and not that good in the low tail. A reason for the latter problems is that the foundations of a sensible selection collapse in low tail, which is packed with national-oriented journals (see for example Zitt et al. 2003). Their impact levels, shaped by confined and incommensurable citation markets, have little significance, jeopardizing the bibliometric part of the selection process.

Option 2

These ways are surely Web 2.0 compatible, but the author should recommend the same rigor and quality control for these processes of "fuzzy peer review" as he rightly advocates for the JIF.

Option 3

JKV will gather a general agreement on his defense of elaborate forms of control and certification. Only the idea that this responsibility can fall on secondary producers is strange, unless Thomson Reuters and perhaps Elsevier-Scopus enter an aggressive strategy of quality control on journals publishing companies.

Variety of science and normalization of impact

JKV mentions normalization and especially field-normalization as one important chapter of the discussion. However, as the main focus of his article is the pathway towards certification of quality measures for journals, the section on normalization is not a strong one.

The JIF is a gross measure, not meant for comparing journals across disciplines. Doing so is, among the misuses, the second mortal sin, that greatest sinners might combine with

[14] The selection process used to be criticized for a bias against non-English speaking journals, European journals, and emerging journals. The situation seems to have improved, due to competition perhaps, and also to the fact that European and non-mainstream journals of good quality turned to English and international openness. Another limitation, quite difficult to cope with, is the possible lack of coverage of small fields in applied science, often importers of knowledge and not well seen from other specialties.

the first one, individual-level evaluation. JIF producers could only be blamed for not posting solemn warnings about it. Moreover, the JCR, together with the JIF and data on citation transactions, lists journals by the SCI classification that immediately suggests a rough normalization. Plenty of literature soon attracted attention to the point and proposed a variety of solutions, cardinal or ordinal, often taking SCI "subject categories" as bases of normalization. In fact, no unique and perfect solution exists: normalization is a profound issue, rooted in the variety of scientific practices and networks, whatever the level of observation.

A ubiquitous issue in scientometrics is the variability of publication practices across fields. It has drawn the attention of scholars since the beginning of evaluative bibliometrics (Narin 1976; Murugesan and Moravcsik 1978; see also Schubert and Braun 1986 *op.cit.*). The contrast between fundamental biology and mathematics, for example, is spectacular. With some generalization and caricature: in mathematics, the number of publications by author is small; the specialization is higher; the collaboration level is typically low, with outstanding exceptions (Erdös); the number of bibliographic references by paper is small; few citations are perfunctory; knowledge obsolescence is mostly low, with frequent late citations. Furthermore, possibly as a result of the cognitive structure of the field, citations are hardly interpretable: high variance of citation scores among fields medalists, without correlation to the peer-judgment about the importance of research. Publication behavior in fundamental biology by and large shows opposite features to mathematics.

The comparison between disciplines or specialties is then hazardous. Two possible responses: either one discipline produces apples and the other oranges, only within-discipline comparisons make sense; or some disciplines produce big apples, other small apples and a careful statistical correction makes sense. Both responses seemingly assume a definition/delineation of disciplines. The comparison across fields is obviously sounder for disciplines obeying the same communication mode. It is easier to compare biology and physics that both privilege articles, than biology and law, since in the latter the articles don't prevail. Other types of differentiation play a role within a field, for example the type of research (applied, theoretical, etc.,).

As every application of citation measures, the impact factor faces the problem of field inequalities, even though the competition for access to better sources has shaped a very unequal market: the main fraction of variability in journal impact, at least for levels of current classifications, is within-field, but the across-fields component is largely sufficient to prevent any sensible inter-field comparison without prior correction. Moreover, as we will see, the notion of "field" can be made superfluous.

Classical normalization: cited-side or target normalization

The principle behind cited-side normalization is "all fields-as targets of citations-are equal". The acceptation of the "normalization" concept is usually more general than a transformation leading to normal distributions. For example the ratio to the field average[15] is often used, under various names: relative impact, mean citation rate, etc. As already mentioned, normalization may also take on other aspects, such as the "relative citation ratios". Rank statistics are also widely used. Literature on normalization is quite abundant (see Sen 1992; Marshakova-Shaikevich 1996; Czapski 1997; Vinkler 2002; Rousseau and

[15] For an actor or a journal i in a field J, the quotient of the actors' impact to the reference's (e.g., world) impact.

Egghe 2003). A general framework for citation indicators normalization not limited to the field issue is found in Raddichi et al. (2008).

Recently, classical indicators, the relative impact on the one hand, the field-level normalization known as "crown indicator" on the other hand, have faced harsh criticism. Lundberg (2007) suggested to normalize first at the article level and then averaging, an argument restated by Leydesdorff and Opthof (2010). Ramanana-Rahary et al. (2009) looking at another variant of normalized indicator, showed that the relative impacts (ratio of impact to the field mean of impact) lead to undesirable properties in aggregation over disciplines and mentioned the weighted average of values based on one frozen level of reference as an alternative[16]. The trend is towards implementing more robust "cited-sided" normalization (Waltman et al. 2011).

Whatever the method, the cited-side normalization in its current practice is a drastic process, which indistinctly laminates all sources of variability: the propensity to cite and to cite quickly; the differences in growth rate; the position of exporter/importer of knowledge to/from other fields; lastly, some imbalances of database coverage among disciplines - a rather good point.

This treatment meets theoretical and practical limits. It is contingent to the delineation of disciplines—how subject categories are delineated—and to the granularity—the option of picking the subject categories 200-level rather than say the 10-level or the 30-level. If one fancies picking up the 10% more cited article in the WoS, the composition of the set will be totally different depending on the normalization level ("cross-scale" or "zoom-level" issue, Zitt et al. 2005): the deeply heterogeneous structure of science, from point of view of citation behavior, partly echoing the self-similar structure of scientific networks suggested by Katz (1999), van Raan (2000) and others, produces unstable results. No satisfactory reference exists, and claims of mathematically optimal level of cut-off, interesting in the neighborhood of some zoom setting, cannot challenge the legitimacy of picking a completely different scale for observing science. Furthermore, most classification schemes result from an unstable mix of political arguments, institutional traditions, and bibliometric considerations.

Alternatives: influence measures

The track of iterative "influence measures" is well known since the pioneer works of Narin and Pinski (*op.cit.*; see also Geller 1978). The recent revival comes from economists (Palacios Huerta and Volij 2004) and biologists (Bergstrom 2007) who implemented at the journal level the "Eigen Factor", integrated to Thomson Reuters journal statistics. Quite a similar approach has been adapted to Scopus by de Moya-Anegon (2007). These implementations allow size-dependent or size-independent measures extremely powerful tools to trace intellectual influences. The relation to normalization is twofold: influence measures "normalize" at the source-level for the prestige of the source, through a recursive process or an equivalent calculation; depending on the implementation, they embody a correction for the propensity to cite, as the citing-side normalization seen below. There is a variety of approaches, depending on the type and granularity of the source considered, the citing journal in the original influence weight, or the citing article or else the citing author. Variants with size-independence properties are proposed. Counterpart of their power, influence measures face some setting problems: treatment of self-citations, timeframe for

[16] This solution, in addition to the relative impacts calculated at each level, was implemented by S. Ramanana-Rahary at OST.

fine grain applications (outside journals), multiple options, and relative complexity of interpretation.

Influence measures are well known and documented in literature. Here we will rather turn towards a less ambitious but promising solution, citing-side or source-level normalization.

Citing-side normalization: logical, classification-free, multidisciplinarity-friendly

The scrutiny of the determinants of the JIF carried a new method of field-normalization experimented at the journal-level by Zitt and Small (2008 "citing side", "fractional citation",[17] and "Audience Factor"). Another option, with finer granularity on citing-side, was explored by Moed (2010), with an industrial application on Scopus ("source-level" SNIP); some other contributions are Zitt 2010; Leydesdorff and Opthof (2010)[18] and Glänzel et al. (2011 "a priori").[19] A related suggestion, albeit less general, was put forth by Nicolaisen and Frandsen (2008).

Background: disassembling the JIF clockwork

In a recent proposal (Zitt 2011), we tried to explicit the relation of the JIF to the determinants previously identified. The general framework of the JIF was assumed. A field-level JIF (FIF) was defined as the weighted average of journals' JIF in a field, admitting a classification scheme for the sake of the exercise.

Let us assume, at a first stage, that the field is isolated, not exchanging citations. With some reasonable working hypothesis, the average field impact factor reveals proportional to:

- The propensity to cite, measured by the length of bibliography of citing articles in the field; and more precisely to cite rapidly, which determines the proportion of "active" references (falling in the citation time-window) to total references, linked to the immediacy-obsolescence conditions in the field.
- The rate of growth of the field over the citation window.

For isolated fields, the average impact does not depend on the size of the field, in contrast with the maximum impact, depending on the skewness of the within-field distribution.

Lifting the hypothesis of isolation, the field average impact factor depends, in addition, on the position of the field in the imports/exports of citations in across-field transactions.

This model is a new formal illustration of Garfield's contributions. Against a widespread opinion he stood firm, attributing the field differences in citation scores to the citation habits rather than to the size of fields: "*The key determinants in impact are not the number of authors or articles in the field but, rather, the citation density and the age of literature cited. The average citations per article and the immediacy of citations are the*

[17] "Fractional citations" were applied by Small and Sweeney (1985) to the metrics of co-citation mapping not to be confused with the fractional count of citations to multiple co-authors in a paper. To our best knowledge, the mention of citing-side normalization for impact calculation, not embedded in influence flows, appears in Zitt et al. (2005 *op. cit*).

[18] The only one cited by JKV. This is all the more regrettable that the papers by Leydesdorff on the topic, so far, show a slightly subjective view of the history of the research front.

[19] If those terms citing-side, fractional, source-level, etc., are equivalent in terms of principle, they were often coined along with particular methodological choices.

significant elements. The size of a field, however, will generally increase the number of super-cited papers" (Garfield 2006).

Citing-side measures

The principle is that, for a convenient definition of sources, "all citation sources are equal". This reverses the point of view: the main factor of inequality identified, the propensity to cite and to cite quickly is corrected at the source, the emission of citations, by some weighting of references inversely to the length of active bibliographies in the neighborhood of each citing article.[20] The citing-side approach is basically another way to count citations and is then applicable to any citation analysis including journal impact.

General features of source-normalized citations are:

- Correction (like usual normalized impact factor; unlike straight impact factor) for the propensity to cite…
- … and to cite rapidly. The auto-correction for the time-frame is built-in by taking only "active" references as a basis of normalization. If the citing-side approach does not of course magically change the time-frame, even 2 years "audience factors" are made comparable across disciplines but of course they rely on scarcer information for slow disciplines. Then, by taking a reasonably large window, say 4 or 5 years, this method spares the heavy management of "variable windows" sometimes proposed in the framework of cited-side normalization (see e.g., Vanclay 2009; Vieira and Gomez 2011) with again the arbitrariness of classifications. Furthermore, the perfect is the enemy of the good: the longer the window, the more frequent the accidents in the journals demography.
- No correction for the difference of growth rate within the timeframe among fields: growing areas are favoured through the boost of the citing set in comparison with the citable set like in JIF, and unlike cited-side normalized JIF.
- No correction for citation exchanges between fields: in contrast with cited-side normalization, the aim is not to make all fields "equal", but "comparable" in a precise sense. Citing-side approach reduces the difference between the average impact of mathematics and biology, by controlling for propensity to cite. But the spirit of the method is avoiding normalization for exports–imports of citations or internal growth of the field. In other words, as far as knowledge and citation are reciprocal flows, knowledge exporters are beneficial, in contrast with knowledge importers. Medical research, showing some cognitive dependence towards cell biology, will tend to show a slightly less scoring in average.
- *Classification-free*: Citing-side normalization is free from ex ante nomenclatures, such as databases classification schemes (e.g., Medline Mesh index, Chemical Abstracts - CAS indexes) or journal classifications of the WoS. In addition, the property is free from any grouping process if the normalization takes place at the journal-level; or strictly at the article level. Locally smoothing the citation behavior over some neighborhood is highly recommended, however, (see below). It is noticeable that "field-normalization" is not an adequate term, not only because no field classification

[20] In the original Audience Factor, the journal level is used for both on citing and cited-side. Emitted citations in the citing journal are weighted in inverse proportion of the average length of bibliographies in this journal's articles. Most other developments use a finer granularity on the citing-side.

is assumed, but also because the normalization is multi-scale: the differences across small areas in the field will be accounted for, as well as the differences across fields.
- The classification-free property allows a natural handling of multidisciplinary literature. The ex post normalization methods face the problem of overlapping fields and suppose the definition of a reference set for multidisciplinary paper/journal. Should a multidisciplinary article, in citable position, be compared to the score of its main field? To some weighted average on all its assigned fields? To an aggregate of all multidisciplinary research, with perhaps a risk of over-normalization? To all science? The citing-side normalization avoids the multiplication of references and the arbitrariness of choice, since it does not need any classification on the cited or on the citing-side. A multidisciplinary article, on the cited-side, aggregates normalized citations which depend on the citation habits of each emitting field.
- *Other applications are appealing*: equalizing citations at the source may be quite helpful for analyzing the transaction of citations issued from the neutralization, in the original network, of the most perturbing effect of citation behavior, the local inequalities of propensity to cite or immediacy. Limit case: it follows from the construction that all isolated areas in the network exhibit the same mean of normalized citation scores (in-links).
- Along with its main goal, the field/area normalization, the citing-side normalization handles other sources of variability insofar as they are expressed in the propensity to cite, which to some extent may be the case for basic and applied research in a same field.

A challenge of citing-side approach is the fine tuning of the normalization and especially the definition of the neighborhood: the smoothing options. The absence of smoothing (each article weighted after its own bibliography), adopted by several authors cited above, is a nice exercise: out-links in the citation network are simply weighted by the inverse of their node degree. However, this choice leads to quite undesirable effects, even mitigated by robust central measures in aggregation. One should avoid the trap of "bad reasons for scarce bibliographies" that would lead to over-rate references coming from marginal articles or articles in trade journals, even when using robust measures in the final cited-level aggregation.

Conclusion: Occam's razor may hurt

Some remarks to conclude:

Despite the huge amount of critical matter accumulated in literature, the JIF still resists, why? Probably through a magic mix of parsimony principle and market power. Often powerful in theoretical matters, the principle of parsimony may be as dangerous as successful in practical ones: Occam's razor may hurt. If we were to pick the most successful indicators in scientometrics, we might come up with the impact factor, the h-index and the ARWU (Shanghai) ranking. All these indicators are both remarkably parsimonious in their framework, quite vulnerable to criticism, and rather seductive to users: getting close to the one-figure indicator is a dream for policy makers and a nightmare for the victims of misuses. In addition, the first two indicators have given birth to a numerous descent of derived products trying to limit the adverse effects at the expense of more complexity, extra settings, making them less attractive to users. The most innovative variant of the JIF,

"influence weights", above-mentioned probably suffered from being more complex and less user-friendly, before its current revival. To face the field-variability, a variety of normalization methods were also published, and new proposals keep on blooming.

The debate is not mature, in several respects:

- The impact factor is outdated and a bit flawed, but the bibliometric or non-bibliometric evaluation of journal impact, along with other citation studies, remains an important and timely issue. Alternatives are multiplying, ranging from cosmetic treatment patching the flaws to radical novelties. In the midst of this abundance of alternatives, sometimes in a process of maturation, one cannot bet that some Occam Razor 2.0, just as sharp but less dangerous than the original, will durably succeed.
- Among unsettled topics, citation normalization, especially field-level normalization, has inspired several proposals too. Citing-side processes have a broad scope of applications, convey a new perspective on the citation network and are especially suited for analyses of multidisciplinarity. Influence measures, with a global view on circulation of knowledge, have also been renewed recently. Another promising area is the development of sophisticated techniques in the wake of utility theory, and more accurate description of indicators properties in the line of axiomatic approaches. They might be applied to journal-level impact, and need some fine-tuning.
- *Non-bibliometric alternatives*: The certification procedures which cannot mainly rely on intermediaries have to comply with the same demanding requirements that those applied to bibliometric indicators. JKV rightly mentions some opportunities, but what is needed is an-depth and comparative evaluation of these processes.

The JIF: angel, devil or scapegoat? All three at a time: angel, for a unique historical role and the many avenues opened to scientometrics and other fields; devil, for a few flaws, and a brightness and market power that may have deterred users from looking aside; but also scapegoat, for misuses and abuses.

Acknowledgments The author thanks S. Ramanana-Rahary and E. Bassecoulard for their help.

References

Adams, J., Gurney, K., & Marshall, S. (2007). Profiling citation impact: a new methodology. *Scientometrics, 72*(2), 325–344.

Adler, R., Ewing, J., & Taylor, P. (2008). Citation statistics. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the International Council of Institute of Mathematical Statistics (IMS). Summarized in (2009). *Statistical Science, 24*(1), 1–14.

Bar-Ilan, J. (2008). Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2), 257–271.

Bergstrom, C. (2007). Eigenfactor: measuring the value and prestige of scholarly journals. *College & Research Libraries News, 68*, 5, www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/may2007/eigenfactor.cfm.

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE, 4*(6), e6022. doi:10.1371/journal.pone.0006022.

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information, 14*(6), 19–47.

Bouyssou, D., & Marchant, T. (2011). Bibliometric rankings of journals based on impact factors: an axiomatic approach. *Journal of Informetrics, 5*(1), 75–86. doi:10.1016/j.joi.2010.09.001.

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics, 69*, 169–173.

Callon, M., & Latour, B. (1981). Unscrewing the big leviathan: how actors macrostructure reality and how sociologists help them to do so. In Karin. D. Knorr Cetina & Aaraon. V. Cicourel (Eds.), *Advances in social theory and methodology: toward an integration of micro- and macro-sociologies* (pp. 277–303). Boston: Routledge and Kegan Paul.

Cronin, B. (1984). *The citation process; the role and significance of citations in scientific communication*. London: Taylor Graham.

Czapski, G. (1997). The use of deciles of the citation impact to evaluate different fields of research in Israel. *Scientometrics, 40*(3), 437–443.

de Moya-Anegon, F. (2007). SCImago. SJR — SCImago Journal & Country Rank.

de Solla Price, D. J. (1963). *Little science, big science*. New York: Columbia University Press.

Garfield, E. (1955). Citation Indexes for Science. A new dimension in documentation through association of ideas. *Science, 122*, 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*, 90–93.

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

Geller, N. L. (1978). Citation influence methodology of Pinski and Narin. *Information Processing and Management, 14*, 93–95.

Glänzel, W. (2008). On some new bibliometric applications of statistics related to the h-index. *Scientometrics, 77*(1), 187–196.

Glanzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics, 53*(2), 171–193.

Glänzel, W., Schubert, A., Thijs, B., & Debackere, K. (2011). A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. *Scientometrics, 87*(2), 415–424.

Hagström, W. O. (1965). *The scientific community*. New York: Basic Books.

Hicks, D. (2004). The four literatures of social science. In H. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research*. New York: Kluwer Academic.

Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences, 104*(49), 19193–19198.

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

Hönekopp, J., & Khan, J. (2011). Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics, 90*(3), 843–853.

Ingwersen, P., Larsen, B., Rousseau, R., & Russell, J. (2001). The publication-citation matrix and its derived quantities. *Chinese Science Bulletin, 46*(6), 524–528.

Katz, S. J. (1999). The self-similar science system. *Research policy, 28*(5), 501–517.

Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology, 61*(11), 2365–2396.

Lundberg, J. (2007). Lifting the crown: citation *z*-score. *Journal of Informetrics, 1*(2), 145–154.

Luukkonen, T. (1997). Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics, 38*(1), 27–37.

Marchant, T. (2009). An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors. *Scientometrics, 80*(2), 325–342.

Marshakova-Shaikevich, I. (1996). The standard impact factor as an evaluation tool of science fields and scientific journals. *Scientometrics, 35*(2), 283–290.

Merton, R.K. (1942). Science and technology in a democratic order. *Journal of legal and political sociology, 1,* 115–126 (reprint: The normative structure of science (1973). In Storer N.W. (ed.), The sociology of science: theoretical and empirical investigations (pp. 1267–1278). Chicago: University of Chicago Press).

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277.

Moed, H. F., & van Leeuwen, T. N. (1995). Improving the accuracy of institute for scientific information's journal impact factors. *Journal of the American Society for Information Science, 46*(6), 461–467.

Moed, H. F., & Vriens, M. (1989). Possible inaccuracies occurring in citation analysis. *Journal of Information Science, 15*, 95–107.

Murugesan, P., & Moravcsik, M. J. (1978). Variation of the nature of citation measures with journal and scientific specialties. *Journal of the American Society for Information Science, 29*(3), 141–155.

Narin, F. (1976). *Evaluative bibliometrics : the use of publication and citation analysis in the evaluation of scientific activity (Report prepared for the National Science Foundation, Contract NSF C-627)*. Cherry Hill: Computer Horizons.

Nicolaisen, J., & Frandsen, T. F. (2008). The reference return ratio. *Journal of Informetrics, 2*(2), 128–135. doi:10.1016/j.joi.2007.12.001.

Palacios Huerta, I., & Volij, O. (2004). The Measurement of intellectual influence. *Econometrica, 72*(3), 963–977.

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

Raddichi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: towards an objective measure of citation impact. *Proceedings of the National Academy of Sciences, 105*(45), 17268–17272.

Ramanana-Rahary, S., Zitt, M., & Rousseau, R. (2009). Aggregation properties of relative impact and other classical indicators: convexity issues and the Yule-Simpson paradox. *Scientometrics, 79*(1–2), 311–327.

Rousseau, R. (2008). Woeginger's axiomatisation of the h-index and its relation to the g-index, the h(2)-index and the r2-index. *Journal of Informetrics, 2*(4), 335–340.

Rousseau, R., & Egghe, L. (2003). A general framework for relative impact indicators. *Canadian Journal of Information and Library Science, 27*(1), 29–48.

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

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

Sen, B. K. (1992). Documentation Note Normalized Impact Factor. *Journal of Documentation, 48*(3), 318–325.

Small, H., & Sweeney, E. (1985). Clustering the science citation index using co-citations: 1. A comparison of methods. *Scientometrics, 7*(3–6), 391–409.

Van Raan, A. F. J. (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics, 47*(2), 347–362.

van Raan, A. F. J. (2001). Competition amongst scientists for publication status: toward a model of scientific publication and citation distributions. *Scientometrics, 51*(1), 347–357.

Vanclay, J. K. (2009). Bias in the journal impact factor. *Scientometrics, 78*(1), 3–12.

Vanclay, J. K. (2011). Impact Factor*:* outdated artefact or stepping-stone to journal certification? *Scientometrics,*. doi:10.1007/s11192-011-0561-0.

Vieira, E. S., & Gomez, J. A. N. F. (2011). The journal relative impact: an indicator for journal assessment. *Scientometrics, 89*(2), 631–651.

Vinkler, P. (2002). Subfield problems in applying the Garfield (impact) factors in practice. *Scientometrics, 53*(2), 267–279.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: some theoretical considerations. *Journal of Informetrics, 5*, 37–47.

Waltman, L., & van Eck, N. J. (2009). A taxonomy of bibliometric performance indicators based on the property of consistency. *Proceedings of the 12th International Conference on Scientometrics and Informetrics*, 1002–1003.

Wouters, P. (1997). Citation cycles and peer review cycles. *Scientometrics, 38*(1), 39–55.

Zitt, M. (2010). Citing-side normalization of journal impact: a robust variant of the audience factor. *Journal of Informetrics, 4*(3), 392–406.

Zitt, M. (2011). Behind citing-side normalization of citations: some properties of the journal impact factor. *Scientometrics, 89*(1), 329–344.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2003). Correcting glasses help fair comparisons in international science landscape: country indicators as a function of ISI database delineation. *Scientometrics, 56*(2), 259–282.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalisation. *Scientometrics, 63*(2), 373–401.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: the audience factor. *Journal of the American Society for Information Science and Technology, 59*(11), 1856–1860.

Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences, 93*(23), 12709–12716.