UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most-highly cited papers

Leydesdorff, L.

DOI
10.1007/s11192-012-0660-6

Publication date
2012

Document Version
Final published version

Published in
Scientometrics

Link to publication

Citation for published version (APA):
Leydesdorff, L. (2012). Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most-highly cited papers. *Scientometrics*, *92*(2), 355-365. https://doi.org/10.1007/s11192-012-0660-6

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:10 Mar 2023

Scientometrics (2012) 92:355–365
DOI 10.1007/s11192-012-0660-6

Alternatives to the journal impact factor: *I3* and the top-10% (or top-25%?) of the most-highly cited papers

Loet Leydesdorff

Received: 23 January 2012 / Published online: 17 February 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Journal impact factors (*IF*s) can be considered historically as the first attempt to normalize citation distributions by using averages over 2 years. However, it has been recognized that citation distributions vary among fields of science and that one needs to normalize for this. Furthermore, the mean—or any central-tendency statistics—is not a good representation of the citation distribution because these distributions are skewed. Important steps have been taken to solve these two problems during the last few years. First, one can normalize at the article level using the citing audience as the reference set. Second, one can use non-parametric statistics for testing the significance of differences among ratings. A proportion of most-highly cited papers (the top-10% or top-quartile) on the basis of fractional counting of the citations may provide an alternative to the current *IF*. This indicator is intuitively simple, allows for statistical testing, and accords with the state of the art.

Keywords Nonparametric · Source normalization · Citation · Journal · Impact

Introduction

In the lead article of this topical issue entitled "Impact Factor: Outdated artefact or stepping-stone of journal certification?" Jerome K. Vanclay focuses primarily on data errors in the database of Thomson Reuters, but less on the statistics of the impact factor (*IF*) as an indicator. The author mentions that the third decimal is provided unnecessarily (in order to minimize the number of tied places; cf. Garfield 2006) and that citation distributions are highly skewed (Seglen 1992, 1997). However, the possible flaws introduced by using averages of these skewed distributions across the file are not elaborated, and significance of differences between impact factors (*IF*s) or the statistical estimation of error in the measurement do not enter into the discussion.

L. Leydesdorff (✉)
Amsterdam School of Communication Research (ASCoR), University of Amsterdam,
Kloveniersbugwal 48, 1012 CX Amsterdam, The Netherlands
e-mail: loet@leydesdorff.net
URL: http://www.leydesdorff.net

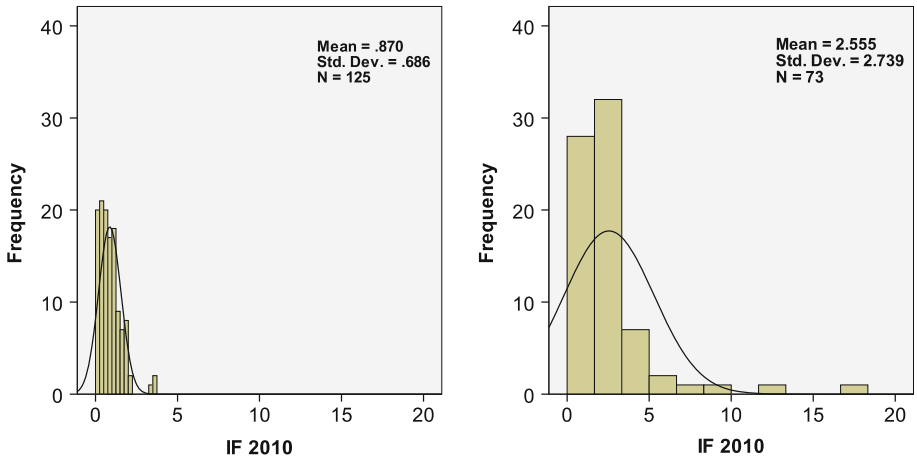

Fig. 1 Impact factors (2010) of 125 journal in the WoS Category "sociology" (XA; *left*) compared with 75 journals in the WoS Category "psychology" (VI; *right*)

The technical problems in the database can increasingly be corrected with further investments in the data processing, but flaws in the data analysis provide an opportunity for scientometric improvement of the indicator. The merit and quality of an indicator depends on its statistical properties and the evaluation of its validity and reliability. In this contribution, I focus on these issues: does the *IF* measure impact? How can one account for differences in citation behavior among fields of science? How can one appreciate the skewness in citation distributions using appropriate statistics?

Stating the problem

Using the same scales, Fig. 1 shows—as an example—the distributions of the *IF*s-2010 of 125 journals classified in the Web of Science under "sociology" (the subject category "XA" in the database) to the left, and the 73 journals classified as "psychology" ("VI") to the right. The two means—0.870 (±0.061) and 2.555 (±0.321), respectively—are significantly different ($p < 0.01$).[1]

Sociology and psychology are neighbouring disciplines. The database additionally distinguishes a category "social psychology" with 56 journals (of which 4 overlap with sociology). The mean *IF*-2010 of this set is 1.499 (±0.169). This distribution is again significantly different from that of sociology journals at the 1% level.[2] Social psychology can be considered as a subfield of psychology, but the distribution of the *IF*s of this subfield is nevertheless significantly different from that of the psychology journals at the 5% level.

Thus, a first problem is that one is not allowed to compare *IF*s even across neighbouring fields and subfields. However, the delineations among fields can also be fuzzy (Boyack and Klavans 2011; Leydesdorff 2006). Scientometricians have used the ISI Subject

[1] *IF*s can be considered as means of citations (in the current year) per publication during the previous two years; therefore, one can expect these mean values to be normally distributed.

[2] The four journals in the overlap were excluded from the comparison of the means. These journals are: *Deviant Behavior*, the *Int. J. of International Relations*, the *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, and *Social Justice Research*..

Categories—recently renamed by Thomson Reuters the WoS Subject Categories—for the normalization, but journals can be attributed to more than a single field, and the attributions themselves are often erroneous (Boyack et al. 2005; Rafols and Leydesdorff 2009). In summary, the problem of the delineation of appropriate sets of journals for the comparison—reference sets—poses a problem that has hitherto remained unresolved.

But even if one accepts that one could compare within these sets—for pragmatic reasons and despite the noted problems—then the normalization in terms of mean values (or *IF*s as 2-year averages) remains unfortunate. Yet, this has been standard practice. How should one then proceed? Schubert and Braun (1986) proposed comparing the mean of the observed citation rates (MOCR) in a sample under study (e.g., during an evaluation) with the mean citation rate in the reference set as the expectation (mean expected citation rate or MECR). These authors introduced the relative citation rate (RCR) as the quotient: RCR = MOCR/MECR.[3] However, the division of two means provides a quotient without a standard error of the measurement (SEM). Consequently, scientometric bar charts and tables often fail to show error bars and to specify uncertainty.

This practice of dividing means was followed by the Center for Science and Technology Studies (CWTS) at Leiden University (Moed et al. 1995) and more recently by the Centre for Research & Development Monitoring (ECOOM) at Leuven University (Glänzel et al. 2009). The relative indicators were renamed with minor modifications as CPP/FCSm in Leiden ("the crown indicator") and NMCR in Leuven. In my opinion, the division of two means contains an error against the order of operations which prescribes first to divide and then to sum. Instead of MOCR/MECR $= \frac{(\sum_{\text{obs}} c_{\text{obs}})/n_{\text{obs}})}{(\sum_{\text{exp}} c_{\text{exp}})/n_{\text{exp}})}$ one should for arithmetical reasons have used the mean of the observed versus expected citation rates, or in formula format: M(OCR/ECR) $= (\sum_{i=1}^{n} \frac{\text{observed}_i}{\text{expected}_i})/n$, in which the expected citation rate is equal to the one derived from the reference set. Unlike MOCR/MECR, M(OCR/ECR) is a normal average with a standard deviation.

This problem was noted by Lundberg (2007), but ignored at the time. Only in 2010 and 2011 did it receive serious discussion in the *Journal of Informetrics* (Opthof and Leydesdorff 2010; Van Raan et al. 2010; Gingras and Larivière 2011). CWTS in Leiden was responsive to the critique and changed the indicator within half a year (Waltman et al. 2011), but the old normalization is still in place in other centers.

What does this discussion mean for the *IF*? Instead of first aggregating the numbers of citations in the current year to citable items in the previous 2 years ($IF = \frac{c_{-1}+c_{-2}}{p_{-1}+p_{-2}}$), one could normalize for citations to each of the previous 2 years separately, as follows: $[(\frac{c_{-1}}{p_{-1}} + \frac{c_{-2}}{p_{-2}})/2]$. The *IF* would then be a moving average with a period of two (Rousseau and Leydesdorff 2011). The difference may be marginal in most cases, but in 2009 the *IF*s of 8.6% of the journals would be changed in the *first* decimal! At the extremes, *Psychological Inquiry* would go from an *IF*-2009 of 4.050 to 9.750 and the *Annual Review of Biophysics* from 19.304 to 9.625.[4] Obviously, statistical decisions matter for the ranking: one can expect the mean of a skewed distribution to be highly sensitive to relatively minor changes in the computation.

[3] One cannot test the *MOCR* against the *MECR* because the two distributions are not independent: the publication set is a subset of the reference set.

[4] In terms of relative decline, the journal *Oceanological and Hydrobiological Studies* would suffer most with a drop of the *IF*-2009 from 0.622 to 0.041.

In summary, in addition to correcting the technical errors in the database as summarized by Vanclay (this issue; cf. Leydesdorff 2008, Table 4 at p. 285) and the arithmetic error in the calculation of these indicators, two scientometric problems remain: (i) how to compare "like with like" (Martin and Irvine 1983; cf. Rafols et al., in press) when the units for the comparison are so different and the differentiations not crisp, and (ii) how to avoid using averages over skewed distributions? In my opinion, important steps towards solutions to these problems have been taken during the last two years.

Comparisons across fields of science

Since the 1980s scientometricians have tried to use the grand matrix of aggregated journal–journal citations for the delineation of fields of science (Doreian and Fararo 1985; Leydesdorff 1986; Tijssen et al. 1987). This matrix can be constructed from the data in the *Journal Citation Reports* (JCR) which have been available (in print) since the mid-1970s. However, the emphasis remained initially on the creation of local journal maps because the decomposition of such a large file (of several thousands of journals) was computationally too intensive for the technology at the time. With the advent of Windows-95 and Pajek in 1996 the decomposition and visualization of large (citation) networks became feasible (Boyack et al. 2005; Leydesdorff 2004). The JCRs are electronically available since 1994.

The conclusion from this research program, in my opinion, has been that any decomposition is beset with error because the sets are not always sufficiently crisp (Leydesdorff 2006). Furthermore, journals themselves are not homogeneous units of analysis in terms of their cognitive contents nor in terms of document types. Letters, for example, have citation half-life times completely different from review articles (Leydesdorff 2008, p. 280). One cannot lump citable items together, and it seems that journals cannot be classified without ambiguity. Classification reduces the data into a tree-like hierarchy, whereas developments take place heterarchically. New entrants (journals), for example, may change the network thoroughly in both cognitively and policy-relevant ways (Leydesdorff et al. 1994).

A solution may be to disaggregate at the level of documents. Documents can be cited in different disciplines and by different types of documents. For example, one can expect papers in the 73 psychology journals used in Fig. 1 to be cited more frequently than papers in the 125 sociology journals. These differences in "citation potentials" (Garfield 1979) can be corrected by "source-normalization" (Moed 2010): the source of the difference is an underlying difference in the citation behavior of the citing authors. More references are expected in some fields than in others. Accordingly, each citation can be fractionally counted, that is, as one over the total number of references (1/NRef) in the citing paper. The field "NRef" is conveniently contained in the WoS database.

Leydesdorff and Opthof (2010) showed that this correction normalizes the huge differences in citation potentials between journals in mathematics to the extent that a leading journal in mathematics (*Annals of Mathematics*) can be ranked even more highly than a major journal in molecular biology despite the latter's (approximately) four times higher *IF* (Table 1). Leydesdorff and Bornmann (2011a) have computed quasi-*IF*s-2008 for 5742 journals in the Science Citation Index (available at http://www.leydesdorff.net/weighted_if/weighted_if.xls). Using regression analysis, they showed that 81% of the variance between 13 fields of science as distinguished in the *Science and Engineering Indicators* of the US National Science Board (2010) is thus corrected, and the remaining differences among these fields are statistically non-significant (cf. Radicchi and Castellano 2012).

Table 1 Comparisons between *Annals of Mathematics* and *Molecular Cells* in Scopus and WoS

	Annals of Mathematics	*Molecular Cells*
IF 2007	2.739	13.156
Fractionally counted quasi-*IF*-2007 on the basis of 3 years in Scopus[a]	0.257	0.386
SNIP 2007	4.979	3.696
IF 2008	3.447	12.902
Factionally counted quasi-*IF*-2008 in the Web of Science[b]	1.416	1.143

Sources [a] Leydesdorff and Opthof (2010, p. 2367) and [b] Leydesdorff and Bornmann (2011a, p. 222)

In addition to this methodological advantage, a conceptual advantage of using citing papers as the reference set for normalization is the delineation ex post in terms of relevant audiences (Zitt and Small 2008). Classifying an evolving system in terms of ex ante categories can be expected to lead to error because the classification system is then defined historically while the dynamics of science is evolutionary (Leydesdorff 1998, 2002; Rafols et al. in press). Using the metaphor of a research front (Garfield 1972, 1979; Price 1965), one would expect important contributions to be made also at the edges of and in between fields (Abbasi et al., in press; Leydesdorff et al. 1994). Authors can be cited in fields unintentionally because the intellectual organization of the sciences is self-organizing as scholarly discourses at the supra-individual and supra-institutional levels (Leydesdorff and Amsterdamska 1990; Fujigaki 1998).

In an evaluation of the different departments of the Tsinghua University in Beijing, for example, Zhou and Leydesdorff (2011) have shown that fractional counting can correct significantly for disadvantages of departments such as those in the arts and humanities when using scientometric evaluations. The Department of Chinese Language and Literature that has previously been rated at the 19th position among 27 departments, was ranked 2nd after the correction for citation potentials reflecting differences in citation behavior among fields of scholarly discourse.

Skewed distributions and non-parametric statistics

In the case of skewed citation distributions, one should avoid central tendency statistics, but use non-parametric statistics such as percentiles (deciles, quartiles, etc.). Bornmann and Mutz (2011) intervened in the discussion about dividing averages or averaging rates by elaborating on the metrics for the six percentile ranks used in the *Science & Engineering Indicators*: top-1%, top-5%, top-10%, top-25%, top-50%, and bottom-50% (National Science Board 2010). Leydesdorff et al. (2011) elaborated these statistics, and Leydesdorff and Bornmann (2011b) applied a newly defined "Integrated Impact Indicator" (*I3*) to two groups of journals: the set of 65 journals classified in the WoS as Information & Library Science, and the 48 "multidisciplinary" journals, including journals such as *Science, Nature,* and *PNAS*.

Figure 2 shows the problem. Using the reference set of 65 Information and Library Science journals, *JASIST* has higher values in *all* six classes, but its *IF*-2009 was only half the size of that of *MIS Quarterly*. Not only is the tale of less-frequently-cited papers in *JASIST* much larger (N of publications = 375), but the 66 most-cited papers in *JASIST*

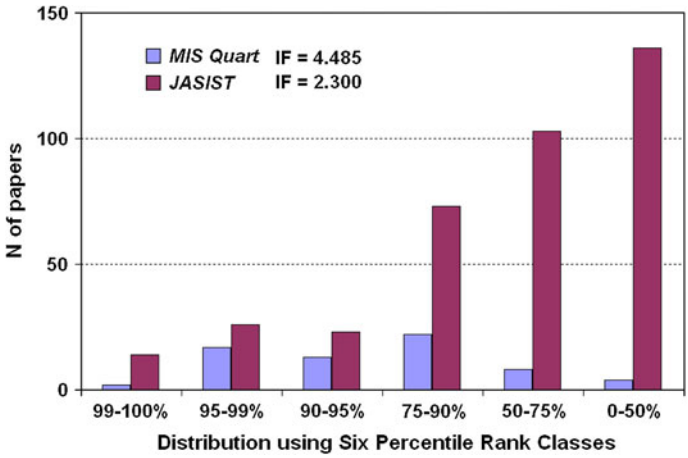

Fig. 2 Six percentile rank classes and *IF*s for *JASIST* and *MIS Quarterly* in 2009 (*source* Leydesdorff and Bornmann, 2011b, p. 2135)

2007 and 2008 are also significantly more cited than the 66 papers in the denominator of the *IF*-2009 of *MIS Quarterly*. Thus, the *IF*s erroneously give the impression that *MIS Quarterly* has an impact higher than *JASIST* (or *Scientometrics* in this set) although in fact it does not.

The misunderstanding is generated by the semantics: the words "impact" and "impact factor" or average impact have been used without sufficient distinction. An average value is determined not only by the numerators, but also the denominators. When less-cited papers are added to a set of highly-cited papers *ceteris paribus*, the total impact of these papers can be expected to increase, but the average impact may decrease. For example, the team of a leading scientist (including postdocs and Ph.D. students) will have more impact than a scientist working alone, but the team's *average* impact is lower.

Accordingly, Bensman (1996) could show that Total Citations—the numerator of the *IF*—could be validated by faculty significantly more than "impact factors" in the evaluation of journals. However, "total citations" are a crude measure. When publications are qualified in terms of percentiles of the citation distribution—instead of by averaging—an integrated impact indicator (*I3*) can be defined as follows:

$$I3 = \sum_i x_i * n(x_i) \tag{1}$$

In this formula, x_i denotes the percentile (rank) value i, and n the number of papers with this value. Instead of averaging, the citation curves are thus integrated after proper normalization to the same scales of hundred percentiles.[5] The scaling makes the distributions comparable. One can also aggregate the percentile values into a normative evaluation scheme such the six classes used in the *US Science and Engineering Indicators* (see Fig. 2).

In Table 2, *MIS Quarterly* and *JASIST* are compared in terms of their *IF*s 2009 and the new indicators, and *Scientometrics* is added to the comparison for the purpose of this discussion. The first and seventh place are precisely reversed between *MIS Quarterly* and *JASIST*, and *Scientometrics* moves from the 10th place in the ranking of *IF*s 2009 to the

[5] The hundred percentiles can be considered as a continuous random variable or more precisely as "quantiles".

Table 2 *MIS Quarterly*, *JASIST*, and *Scientometrics* compared in terms of total citations, *IF*-2009, *%I3*, and %PR6 within the reference set of 65 LIS journals

Journal	N of papers	Total citations	*IF* 2009	% *I3*	% PR6 (six ranks)
MIS Quart	66	847	4.485 [1]	2.61 [7]$^{+}$	2.34 [7]$^{+}$
J Am Soc Inf Sci Technol	375	1,975	2.300 [7]	9.73 [1]$^{+}$	8.63 [1]$^{+}$
Scientometrics	258	1,336	2.167 [10]	7.24 [2]$^{+}$	6.37 [2]$^{+}$

Source Leydesdorff and Bornmann (2011b, p. 2139)

$^{+}$ $p < 0.01$; above the expectation. Ranks are added between *brackets*

second place using both *I3* (on the basis of quantiles) and the six percentile ranks (*6PR*) as indicators. All three journals are cited above expectation given the reference group of 65 LIS journals ($p < 0.01$).

The integrated impact indicator (*I3*) thus combines a number of advantages:

1. *I3* values can be tested against expectations using the z test for two independent proportions.
2. *I3* values are determined at the article level. A journal can thus be defined as one possible set of papers, but other aggregations remain equally possible. For example, one can also compare countries or institutions in terms of their *I3* values (Leydesdorff 2011).
3. Percentiles can be aggregated differently in terms of the (normative) evaluation scheme chosen in a given policy context. I mentioned the evaluation scheme of six ranks used by the US-NSF, but the research assessment exercises (RAE) in the UK, for example, has hitherto used 4 + classes (e.g., Rafols et al. in press).
4. Rousseau (in press) noted that the popular indicator of the top-10% most-highly-cited papers (Tijssen et al. 2002)—e.g., the *Excellence Indicator* of the new edition of the SCImago Institutions Rankings, and also used for the Leiden Ranking 2011/2012 (CWTS 2011)—can be considered as a special case of two percentile rank classes. Bornmann et al. (in press) elaborated the test statistics for this special case.
5. *I3* values can be used across databases; for example, the user may wish to include "grey literature" or so-called non-source references (in the WoS) in the reference set (e.g., Bensman and Leydesdorff 2009). However, the definition of a reference set remains a requirement (Rousseau, in press). In my opinion, this limitation makes the analyst reflexively aware that each set is a sample and that impact values are sample-dependent.
6. *I3* values correlate both with the number of publications and with the numbers of citations because they are based on the (scalar) sum of the multiplications of these two numbers. Citations themselves can be considered as impact indicators, and publications as performance indicators; they may correlate because of scale effects. In the cases that we tested, the correlations between *I3* and total citations or total publications were higher than the correlations between these latter two (Table 3).

In sum, *I3* provides a measure that is statistically warranted and leaves the user free to select from a number of options, such as the choice of a normative evaluation scheme. One can also test heterogeneous sets, such as departments in a university or projects within a program, against one another. The problems with the statistics involved in measuring

Table 3 Rank-order correlations (Spearman's ρ; upper triangle) and Pearson correlations r (lower triangle) for the 48 journals in the "multidisciplinary" set of the WoS

Indicator	*IF*-2009	*I3*	Number of publications	Total citations
IF-2009		0.798**	0.479**	0.840**
I3	0.590**		0.829**	0.986**
N of publications	0.492**	0.953**		0.772**
Total citations	0.841**	0.922**	0.839**	

Source Leydesdorff and Bornmann (2011b, p. 2142)

** Correlation is significant at the 0.01 level (2-tailed); * correlation is significant at the 0.05 level (2-tailed)

impact can thus be solved. *I3* can be used both for whole-number counted and fractionally-counted citation rates.

Next steps

Hitherto, we have not combined the two proposals, but studied *I3* in journals belonging to a single WoS Category (Leydesdorff and Bornmann 2011b) or specified subsets thereof (Leydesdorff 2011). If the unit of analysis for an evaluation, however, is multi-disciplinary such as in the case of a university, one can combine the two normalizations and use *I3*-values based on fractional citation scores.

At the level of the WoS or Scopus databases—which are multi-disciplinary in nature—the fractionalization of the citation counts would take care of the differences in "citation potentials" (Garfield 1979) both synchronically as diachronically (Althouse et al. 2009; Radicchi and Castellano 2012, p. 129) without imposing a priori categorization of journals in subject categories. In an email communication (23 June 2010), Ludo Waltman suggested that a remaining difference among fields of sciences might be caused by the different rates at which papers in the last 2 years are cited in different fields. Correction for this effect would require one additional normalization at the level of each *journal*.

The further introduction of non-parametric statistics into the system may take more time because of existing institutional routines. Most recently, however, both the SCImago Institutions Rankings (at http://www.scimagoir.com/pdf/sir_2011_world_report.pdf) and the Leiden Ranking 2011/2012 (at http://www.leidenranking.com/ranking.aspx) introduced the 10% most-highly cited papers as the *Excellence Index* and *Proportion top-10% publications* ($PP_{top\ 10\%}$), respectively. Bornmann and Leydesdorff (2011) used this same standard for overlays on Google Maps. These excellence indicators for the Scopus and WoS databases, respectively, allow for statistical testing of the significance of differences and rankings (Bornmann et al., in press; Leydesdorff and Bornmann, in press). As noted, the top-10% most-highly cited can be considered as a special case of *I3* (Rousseau, in press). This measure thus has all the advantages listed above. It may be easier to understand this measure than *I3* on the basis of quantiles or its equivalent using six percentile ranks (PR6).

Given the increasing consensus about the proportion of the top-10% most-highly cited papers as an excellence indicator, one could also explore this measure as an alternative to the *IF*. Using fractional counting of the citations and with proper normalization for different document types, the differences of "citation potentials" of journals in different fields of science can significantly be reduced (Leydesdorff and Bornmann 2011a). The indicator

is intuitively simple, allows for statistical testing, and accords with the current state of the art.

Further research is needed because the proportion of 10% most-highly cited documents may insufficiently distinguish among a potentially large group of journals with no or few publications in the top-10%. In the case of patent evaluation, Leydesdorff and Bornmann (in press, b), for example, used the top-25% for this reason (cf. Adams et al. 2000). The top-quartile may be more useful than the top-10% in the case of journals, but this issue has to be informed by empirical research. In addition to the excellence indicator, *I3* and/or PR6 provide impact indicators which allow for comparisons among less excellent units of analysis by taking also their productivity into account.

Acknowledgement I am grateful to Lutz Bornmann for comments on a previous draft.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Abassi, A., Hossain, L., & Leydesdorff, L. (in press). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*.

Adams, J., Cooke, N., Law, G., Marshall, S., Mount, D., Smith, D., & Stephenson, J. (2000). *The role of selectivity and the characteristics of excellence. Report to the Higher Education Funding Council for England* (p. 76). Leeds, UK/Philadelphia, PA: Evidence Ltd., Higher Education Policy Unit, University of Leeds, ISI; Retrieved on January 18, 2012 from http://www.evidence.co.uk/downloads/selectivity-report.pdf.

Althouse, B. M., West, J. D., Bergstrom, C. T., & Bergstrom, T. (2009). Differences in impact factor across fields and over time. *Journal of the American Society for Information Science and Technology, 60*(1), 27–34.

Bensman, S. J. (1996). The structure of the library market for scientific journals: The case of chemistry. *Library Resources & Technical Services, 40*, 145–170.

Bensman, S. J., & Leydesdorff, L. (2009). Definition and identification of journals as bibliographic and subject entities: Librarianship vs. ISI journal citation reports (JCR) methods and their effect on citation measures. *Journal of the American Society for Information Science and Technology, 60*(6), 1097–1117.

Bornmann, L., de Moya-Anegón, F., & Leydesdorff, L. (in press). The new excellence indicator in the world report of the SCImago Institutions Rankings 2011. *Journal of Informetrics*.

Bornmann, L., & Leydesdorff, L. (2011). Which cities produce excellent papers worldwide more than can be expected? A new mapping approach—using Google Maps—based on statistical significance testing. *Journal of the American Society for Information Science and Technology, 62*(10), 1954–1962.

Bornmann, L., & Mutz, R. (2011). Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. *Journal of Informetrics, 5*(1), 228–230.

Boyack, K. W., & Klavans, R. (2011). Multiple dimensions of journal specificity: Why journals can't be assigned to disciplines. In E. Noyons, P. Ngulube, & J. Leta (Eds.), *The 13th conference of the international society for scientometrics and informetrics* (Vol. I, pp. 123–133). Durban, South Africa: ISSI, Leiden University and the University of Zululand.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351–374.

CWTS. (2011). *The Leiden ranking 2011/2012: Data collection and indicators*. http://www.leidenranking.com/leidenranking.pdf. Retrieved on Dec 10, 2011.

Doreian, P., & Fararo, T. J. (1985). Structural equivalence in a journal network. *Journal of the American Society for Information Science, 36*, 28–37.

Fujigaki, Y. (1998). Filling the gap between discussions on science and scientists' everyday activities: Applying the autopoiesis system theory to scientific knowledge. *Social Science Information, 37*(1), 5–22.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*(Number 4060), 471–479.
Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359–375.
Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*(1), 90–93.
Gingras, Y., & Larivière, V. (2011). There are neither "king" nor "crown" in scientometrics: Comments on a supposed "alternative" method of normalization. *Journal of Informetrics, 5*(1), 226–227.
Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics, 78*(1), 165–188.
Leydesdorff, L. (1986). The development of frames of references. *Scientometrics, 9*(3–4), 103–125.
Leydesdorff, L. (1998). Theories of citation? *Scientometrics, 43*(1), 5–25.
Leydesdorff, L. (2002). Dynamic and evolutionary updates of classificatory schemes in scientific journal structures. *Journal of the American Society for Information Science and Technology, 53*(12), 987–994.
Leydesdorff, L. (2004). Clusters and maps of science journals based on bi-connected graphs in the journal citation reports. *Journal of Documentation, 60*(4), 371–427.
Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal–journal citation relations using the journal citation reports? *Journal of the American Society for Information Science and Technology, 57*(5), 601–613.
Leydesdorff, L. (2008). *Caveats* for the use of citation indicators in research and journal evaluation. *Journal of the American Society for Information Science and Technology, 59*(2), 278–287.
Leydesdorff, L. (2011). *Comparison of the integrated citation impacts of journals, nations, and institutions in the set journals of "nanoscience & nanotechnology"*. Paper presented at the Atlanta conference on science and innovation policy, Atlanta, GA; September 15–17.
Leydesdorff, L., & Amsterdamska, O. (1990). Dimensions of citation analysis. *Science, Technology and Human Values, 15*(3), 305–335.
Leydesdorff, L., & Bornmann, L. (2011a). How fractional counting affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology, 62*(2), 217–229.
Leydesdorff, L., & Bornmann, L. (2011b). Integrated impact indicators (I3) compared with impact factors (IFs): An alternative design with policy implications. *Journal of the American Society for Information Science and Technology, 62*(7), 1370–1381.
Leydesdorff, L., & Bornmann, L. (in press). Mapping (USPTO) patent data using overlays to Google maps.
Leydesdorff, L., & Bornmann, L. (in press). Testing differences statistically with the Leiden Ranking. *Scientometrics*.
Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables in citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology, 62*(7), 1370–1381.
Leydesdorff, L., Cozzens, S. E., & van den Besselaar, P. (1994). Tracking areas of strategic importance using scientometric journal mappings. *Research Policy, 23*, 217–229.
Leydesdorff, L., & Opthof, T. (2010). *Scopus*' source normalized impact per paper (SNIP) versus the 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.
Martin, B., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*, 61–90.
Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277.
Moed, H. F., De Bruin, R. E., & Van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics, 33*(3), 381–422.
National Science Board. (2010). *Science and engineering indicators*. Washington DC: National Science Foundation. http://www.nsf.gov/statistics/seind10/.
Opthof, T., & Leydesdorff, L. (2010). *Caveats* for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics, 4*(3), 423–430.
Price, D. J. d. S. (1965). Networks of scientific papers. *Science, 149*(no. 3683), 510–515.
Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics, 6*(1), 121–130.
Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. *Journal of the American Society for Information Science and Technology, 60*(9), 1823–1835.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (in press). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*.

Rousseau, R. (in press). Basic properties of both percentile rank scores and the *I3* indicator. *Journal of the American Society for Information Science and Technology*. doi:10.1002/asi.21684.

Rousseau, R., & Leydesdorff, L. (2011). Non-consistency, non-cited items, and the impact factor: A consequence of the arithmetic. *ISSI Newsletter, 7*(1), 10–14.

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

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

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*, 498–502.

Tijssen, R., de Leeuw, J., & van Raan, A. F. J. (1987). Quasi-correspondence analysis on square scientometric transaction matrices. *Scientometrics, 11*(5–6), 347–361.

Tijssen, R., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381–397.

van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J., & Waltman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics, 4*(3), 431–435.

Vanclay, J. K. (in press). Impact factor: Outdated artifact or stepping-stone of journal certification? *Scientometrics*.

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*(1), 37–47.

Zhou, P., & Leydesdorff, L. (2011). Fractional counting of citations in research evaluation: A cross- and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics, 5*(3), 360–368.

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.