

FROM THE EDITOR'S DESK

Impact Factor—a Useful Tool, but Not for All Purposes

Christopher Baethge

How likely is it that a scientific article will still be being cited half a century after it first appeared? A working group around John P. Ioannidis of Stanford University has investigated the fate of more than 5000 contributions in 27 medical journals from the year 1959 (1), and found that no fewer than one in 23 of these texts received a citation in 2009. Classic status, however—at least five citations—was attained by only one in 400 publications; examples given in the study are the descriptions of Prinzmetal's angina or Down syndrome.

However, the study also indicates how ephemeral success is for some periodicals. The three journals with the highest impact factors in 1959 (*Medicine*, *American Journal of Medicine*, and *British Medical Bulletin*) are now in the middle of the field, while four have ceased publication altogether. Another phenomenon that appears is the anglicization of medicine. Of the four German-language journals on the list in the middle of the 20th century, today only the *Deutsche Medizinische Wochenschrift* remains. Above all, however, the analysis produced one interesting result for the discussion of the impact factor: There was a significant correlation (0.87) between the 2-year impact factor usually used today and the 50-year impact factor. In other words, the much criticized most important evaluation measure in science predicted the influence of articles over a period of five decades pretty well.

So is the impact factor, against much current opinion, not so bad a criterion after all for the evaluation of science? With the approach of the publication of the impact factors for 2011 in the middle of this year, and the enormous significance of this event for research department funding, this is a good moment to call to mind one or two facts about the impact factor.

Influence, not quality

First of all, just as was intended by its inventor, the American bibliometrician Eugene Garfield, the impact factor is not an absolute measure of the quality of a journal, but of its influence. Therefore, choosing the impact factor as an evaluation instrument equates to a decision to measure influence and not, or not primarily, quality. True, there should be a correlation between influence and quality, but it is difficult to agree on a definition of scientific quality. This is noticeable, for example, when researchers debate the quality of their

own work and that of colleagues. If one nevertheless wishes to evaluate research, then in the absence of a generally recognized concept of quality, the obvious step is to make use of a surrogate measure.

Among the many surrogate parameters, the impact factor has achieved particular popularity, especially in government research funding and medical schools (2). There is nothing inevitable about this: The main databases—those of Thomson Reuters (Journal Citation Report) or Elsevier (Scopus), for example—also make available more highly elaborated parameters, e.g., the Hirsch factor or Eigenfactor.

The impact factor (IF), which relates the number of citations of a journal to the number of articles it publishes, is used not just to rate journals—the purpose Garfield intended it for—but also to evaluate researchers and institutions. However, while the scientific influence of journals appears to be well reflected in the IF, caution is needed when evaluating authors. Anthony van Raan, of the Centre for Science and Technology Studies of the University of Leiden, believes that “if there is one thing every bibliometrician agrees, it is that you should never use the impact factor to evaluate research for an article or for an individual—that is a mortal sin” (3). Or, in Eugene Garfield's own words (4): “It is one thing to use impact factors to compare journals and quite another to use them to compare authors.”

Distribution of citations

How can something be a good measure of the influence of journals, but a bad measure of the influence of individual researchers? The explanation lies in the distribution of citation frequencies. Even for very influential journals, most citations relate to a small group of articles: Of all the reference citations in 2004 referring to contributions published in *Nature* in 2002 and 2003, about 90% related to only 25% of the articles (5). Most of the contributions in *Nature* were cited fewer than 20 times. An article in the top one-fourth of frequently cited contributions was mentioned on average about 24 times as often as one in the bottom three-fourths. In an evaluation using IF, however, the authors of all the articles in *Nature* are treated the same.

The figures for other journals confirm these data. Out of all the citations in other journals in 2009 of

articles published in *Deutsches Ärzteblatt International*, 65% related to 15% of its scientific contributions.

The author of an article in a specialty journal with a lower IF may therefore be cited just as often or even more often than the author of a contribution in a higher-rated journal.

A working group in the International Mathematical Union has calculated the risk of incorrect ranking of researchers on the basis of the journal impact factor taking the sample case of two mathematics periodicals (6). *Transactions of the American Mathematical Society* has an IF that is roughly twice that of the *Proceedings of the American Mathematical Society*, so a *Transactions* author would have an evaluation advantage over a *Proceedings* author. In actual fact, though, there is a 62% chance that the *Proceedings* article would be cited at least as many (or as few) times. So in six out of ten cases, the evaluation would have led to an unfair ranking. The example is no doubt particularly representative of journals with a low impact factor, and one would hope that such effects would average out when many publications by many authors are compared. Nevertheless, it illustrates the danger lurking in careless use of the journal impact factor.

Alternatives

It must, however, be considered what the consequences would be of abandoning the journal impact factor in favor of taking more account of author citations, as already practiced with the Hirsch factor (which corresponds to the number x of articles by an author that have been cited at least x times). This could lead to the development of a citation behavior in which considerations of friendship or politics play an even larger part than is already the case. This is why the German Research Foundation (Deutsche Forschungsgemeinschaft) is trying to abandon quantitative measures altogether and has decided to base its evaluation of research applications only on a small number of important publications of its applicants, partly in order to make it possible for the peer reviewers to actually read the texts and appreciate their content (7). Whether it is possible to evaluate these articles independently of knowledge about impact factors remains an open question. Above all, one cannot rule out the possibility that this route will lead to the reappearance of exactly what the supporters of objective evaluation criteria such as the impact factor want to prevent: decisions made on the basis of subjective or even random criteria.

Major differences between specialties

One thing that is striking about impact factors is the differences between the various specialties. The *Table* lists as an example the impact factors of 27 subject specialties in human medicine, mostly clinical. Even at this level the median IFs differ by a factor of 2.6. The gradient becomes even more marked between the impact factors of the highest-rated journals of different specialties. The IF of the best-placed oncology journal

TABLE

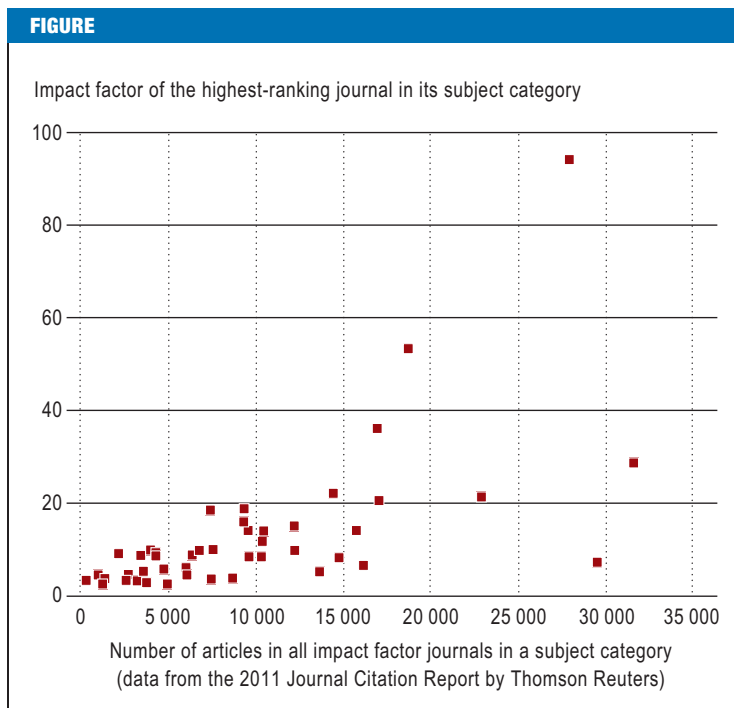
Median and highest impact factor (IF) in clinical specialties

Category	Median IF of all journals	Δ vs. 2005 (%)	IF of the top journal	Δ vs. 2005 (%)
Endocrinology & metabolism	2.796	+22.1	22.469	-0.3
Hematology	2.747	+36.8	14.432	+24.1
Infectious diseases	2.594	+9.4	16.144	+61.3
Oncology	2.455	+3.5	94.333	+89.4
Pharmacology & pharmacy	2.355	+24.7	28.712	+44.8
Critical care medicine	2.353	+69.8	10.191	+17.3
Respiratory system	2.272	+36.6	10.191	+17.3
Gastroenterology & hepatology	2.21	+22.5	12.032	-2.9
Anesthesiology	2.176	+43.9	5.486	+27.3
Geriatrics & gerontology	2.029	+1.2	9	+5
Psychiatry	2.011	-1.7	15.47	+22.4
Clinical neurology	1.994	+23.1	21.659	+92.9
Cardiac and cardiovascular systems	1.993	+27.8	14.432	+24.1
Pathology	1.913	+17.8	18.778	+202.2
Radiology & nuclear medicine	1.861	+19.1	7.022	+30.6
Urology & nephrology	1.843	+19.4	8.843	+22.1
Dermatology	1.667	+27.1	6.27	+42.3
Obstetrics & gynecology	1.616	+10.1	8.755	+60.7
Ophthalmology	1.362	+7.8	10.34	+36.5
Dentistry, oral surgery and medicine	1.345	+4.4	3.933	0
Pediatrics	1.314	+19.2	5.391	+26.2
Emergency medicine	1.269	+38.1	4.177	+49
Surgery	1.236	+21.5	7.474	+18.1
Orthopedics	1.164	+14.3	3.953	-6.2
Medicine, Legal	1.159	+13	2.939	+34.1
Medicine, General and Internal	1.104	+13.7	53.486	+21.5
Otorhinolaryngology	1.067	+19.2	3.038	+15
Average	1.848	+20.9	15.517	+36.1

Data are the current impact factors (impact factors for 2010, published in 2011) of 27 selected medical categories.

Δ, Change from the corresponding value for 2005.

The ratios of the largest values to the smallest values in the two impact factor columns are 2.6 (median IF) and 31.1 (highest IF).



(*CA: A Cancer Journal for Clinicians*) is 30 times as high as the corresponding figure in forensic medicine (the IF of the *International Journal of Legal Medicine*).

Compared to 2005, the median and highest impact factors in subject specialties have almost all risen, with the journals at the top of the impact factor list having gained more than the crowd. This rise itself continues the growth of the impact factors that has been observed for years. In 1982 the *New England Journal of Medicine* had an IF of 11.4 (today it is 53.5) and was thus just behind the *Lancet* with its 11.6 (33.6 today). One main reason for the impact factor boom is the growth in research production: More and more researchers are writing more and more contributions in more and more sources (8, 9) with more and more references. Of note, there is a correlation between IF and the quantity of articles in a given research field: For the impact factor of the highest-rated journal in each of 46 selected medical subspecialties (excluding basic subjects) and the number of articles published in these fields, the correlation coefficient is 0.67 (Spearman, $p < 0.0001$) (Figure). The correlation drops when instead of the IF of the best-placed title one uses the median IF and hence looks at the midfield journals: $r = 0.25$;

$p = 0.095$; difference between the correlations: $p = 0.01$. This is further confirmation of the Matthew principle (“To him that hath, more shall be given”), visible here as the particular benefit gained by high-ranking journals from the growth in the medical sciences. Above all, however, the numbers may indicate that the size of a research field plays a part in the size of the impact factor. This is why many bibliometricians, including those who produce the impact factor itself, recommend using the impact factor with caution, e.g., through the use of field-adjusted impact factors (10, 11).

Conflict of interest statement

The author is Chief Scientific Editor of *Deutsches Ärzteblatt* and *Deutsches Ärzteblatt International*.

Translated from the original German by Kersti Wagstaff, MA.

REFERENCES

- Ioannidis JPA, Belbasis L, Evangelou E: Fifty-year fate and impact of general medical journals. *PLoS ONE* 2010; 5(9): e12531.
- Brähler E, Strauß B: Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten. Eine aktuelle Übersicht. *Bundesgesundheitsbl* 2009; 9: 910–6.
- v. Noorden R: A profusion of measures. *Nature* 2010; 465: 864–6.
- Garfield E: The agony and the ecstasy – the history and the meaning of the journal impact factor. Lecture given on 16 September 2005; <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>
- Nature Editors: Not so deep impact. *Nature* 2005; 435: 1003–4.
- Adler R, Ewing J, Taylor P: Citation statistics. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). 2008; www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf
- Kleiner M: Qualität statt Quantität – Neue Regeln für Publikationsangaben in Förderanträgen und Abschlussberichten. Press conference on 23 February 2010; www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2010/statement_qualitaetstatt_quantitaet_mk_100223.pdf
- Wuchty S, Jones BF, Uzzi B: The increasing dominance of teams in production of knowledge. *Science* 2007; 316: 1036–9.
- Mabe M: The growth and number of journals. *Serials* 2003; 16: 191–7.
- AWMF: AWMF-Vorschlag zur Verwendung des Impact-Faktors. 2001; www.awmf.org/forschung-lehre/kommission-fl/forschungsevaluation/bibliometrie/impact-faktoren.html
- Thomson Reuters: The Thomson Reuters Impact-Factor. http://thomsonreuters.com/products_services/science/free/essays/impact_factor/ (revised 1994 article by Eugene Garfield).

Corresponding author

Prof. Dr. med. Christopher Baethge
 Leiter der Medizinisch-Wissenschaftlichen Redaktion
 Deutsches Ärzteblatt, Ottostraße 12, 50859 Köln
 baethge@aerzteblatt.de

Cite this as

Baethge C: Impact factor—a useful tool, but not for all purposes. *Dtsch Arztebl Int* 2012; 109(15): 267–9. DOI: 10.3238/arztebl.2012.0267