

What do impact factors tell us?

There is an increasing tendency, when considering appointment, tenure or promotion, or when comparing research groups, to evaluate researchers according to the citation index values (Box 1) of the journals in which they publish. Although the Research Assessment Exercise (a process that determines UK government funding) does not explicitly use impact ratings, some British universities have set a standard of expecting that all biologists publish in journals with impact factors of at least five and that those with ratings below four are likely to be excluded from the exercise. A concern is that such a criterion is unfair¹⁻³, especially in comparisons among fields of research.

Here, we show how this usually disadvantages evolutionary biologists and ecologists. One widespread belief is that the size of the field (measured as the annual number of published articles) largely determines the impact^{3,4}, suggesting a need for rescaling. We show that this proposition is both logically incorrect and unsupported by data, but that various other artefacts also influence journal impacts. We use a population demography approach to derive the effect of factors such as the average number of references per paper, fraction of references that are in journals included in the Science Citation Index, growth of the subject or journal, and interval between publication and reference. We show that directions of change are not always clear: growth of the subject as a whole can lead to either increasing or decreasing impacts, depending on the time lag between publication and citation. Most importantly, we refute the claim⁵ that comparisons of impacts between fields can be used to assess research quality and show that within a clearly defined field, judgement is only possible if there is no appreciable subdivision of research practices.

A simple model of citation bookkeeping (Box 2) shows that differences in impact factors among fields cannot logically depend on any aspect of research 'quality' at all. Because of competition among citations, papers in a field twice the 'quality' of another would still have a fixed number of references, which focus on the best papers within that field, and neglect the (relatively) less interesting ones. In a field where all science is poor, papers are still being cited – and if the number of references per paper equals that of the 'good science' field, the impact factors must be exactly same.

Within a field, the average number of citations per paper depends simply on the average number of references per paper. Variation in this mean (along with a series of sampling issues) then generates variations between fields (Box 2).

Only papers published in journals included in the Science Citation index are included in impact calculations. Books, chapters in books and nonlisted journals are thus excluded, so reducing the number of citable references per paper and thus reducing the impact. Consequently, subjects whose papers regularly quote books or include information from obscure sources, for example to give details of study sites or details of the basic ecology of the study species, will have a reduced impact.

Critically, only very recent citations are included in citation indices so that fields with a high impact factor tend to be those with a short half-life of citations (Fig. 1 and Box 2). A short half-life results in a higher proportion of citations before the deadline of a few (usually two) years in which citations are counted (Eqn 1 in Box 2). Time differences in citation accumulation⁶ can result from several origins. In some subjects, which focus on few very recently developed questions at one time, research can be carried out and published quickly. In other subjects, research progresses on a larger range of questions, with each question requiring long-term effort. A very active field in which each paper inspires a new research project would obtain zero impact, if the completion to publication of each project takes more time than the impact time window (Box 2).

As an illustration of this difference, the highest ranking journal in mathematics has an impact factor equivalent to the 51st in cell biology⁷. This low performance clearly does not indicate that the quality of mathematics as a science is poor. Instead, it probably arises from the tradition of mathematical research progressing on many questions simultaneously, where papers quote few references, but where each paper contains a large amount of relevant information and has a long publication time. Consequently, published information is useful for a long time.

This is exemplified by the proof of Fermat's Last Theorem for which the original reference is a scribble in the margin of Diophantus' *Arithmetica* (the margin text was published in 1670, after Fermat's death), and the paper⁸ that provides the solution quotes 84 references of which only four are published in the previous two years – hardly surprising as the refereeing took a year and the author, Andrew Wiles, had previously devoted seven years to the problem⁹. Indeed, the ten mathematical journals with the highest impact factors all have half-lives of at least 9.9 years, whereas none of the same subset of cell biology journals achieves a half-life of six years⁷.

The number of papers published within a field does not affect the average citations at all (Box 2) and hence has nothing to do with the impact factor of a field¹⁰. The size-independence can be understood by imagining an isolated subject with only one journal, which therefore always quotes itself. If each paper in this journal had 20 references to work published in the same journal in the previous year, the impact factor of this journal would equal 20 (assuming a time window of one year), regardless of the

Box 1. Some widely used terms, and their definitions

Impact factor of a journal: the average number of times that articles published in a specific journal in the two previous years (e.g. 1997–1998) were cited in a particular year (e.g. 1999). The citing journals have to be included in the Science Citation Index.

Citation index of a journal: same as impact factor.

Impact time window: number of years included in the calculation of the impact factor (usually two, as indicated above).

Impact factor of a field: the average number of times that articles published in journals of that field in the two previous years (e.g. 1997–1998) were cited in a particular year (e.g. 1999). This can be calculated by weighting journal impact factors by the number of articles they have published. The citing journals have to be included in the Science Citation Index.

Half-life of a journal: the number of years, going back from the current year, that account for 50% of the total citations received by the journal in the current year. If 1000 citations appear in year 1999 that refer to journal X and 500 of them refer to papers published in 1996 or later, the half-life of journal X equals three years. A long half-life implies that a small proportion of citations are included in the time window that influences impact factors.

Half-life of a field: the number of years, going back from the current year, that account for 50% of the total citations received by journals within that field in the current year.

Growth rate of a field or a journal: the number of papers appearing in a specific year divided by the number in the preceding year.

Box 2. A demographic model of impact factors

Assume that all science is divided into F different ‘fields’, such as astronomy or virology, among which impact factors are to be compared. A field i has produced $N_i(t)$ scientific articles t years ago ($t = 0$ marks current year), and the number of references in each of these papers averages $c_i(t)$. The citation index of journal j which produces $n_j(t)$ papers, can then be written as follows

$$CI_j = \frac{\sum_{i=1}^F \sum_{t=1}^T N_i(t) c_i(t) p_{ij}(t, 0)}{\sum_{t=1}^T n_j(t)} \tag{1}$$

assuming the index is calculated using citations within T years (typically $T = 2$). The factor $p_{ij}(t_1, t_2)$ gives the probability that a paper that appeared t_1 years ago in journal j is cited in a randomly chosen article published t_2 years ago ($t_1 \geq t_2$) and belonging to field i . Assuming that the journal j belongs to field $f(j)$, this factor can be further divided into the following components:

$$p_{ij}(t_1, t_2) = \frac{N_{r(f(j))}(t_1) R_{i,r(f(j))}(t_1 - t_2)}{\sum_{\tau=0}^{\infty} \sum_{k=1}^F N_k(\tau) R_{ik}(\tau)} \cdot \frac{n_j(t_1) r_j}{N_{f(j)}(t_1)} \tag{2}$$

$R_{i,r(f(j))}(t)$ is the average relevance of a paper belonging to field $f(j)$ to the development of field i on a time scale of t years, defined such that the numerator of the first term gives the probability that a randomly chosen citation in a paper of field i refers to field $f(j)$ when the respective publication years differ by $t_1 - t_2$. It is clear that $R_{i,r(f(j))}(t)$ typically peaks within the same field, thus when $i = f(j)$, and with a time difference of a few years; also, $R_{i,r(f(j))}(t)$ has a wider distribution in fields with long citation half-lives (Box 1). $R_{i,r(f(j))}$ can hardly function as a measure of research quality – it is not a ‘fault’ of virology if its results are of little relevance for making progress in, say, astronomy. Instead, r_j which gives the relative relevance of results published in journal j compared with the average of its field, reflects the importance of a result for further work: if the relevance r_j of one journal is twice of that of another in the same field, results published in it are cited twice as often. To give correct probabilities in Eqn 2, the relevance values must scale according to $\sum_j n_j(t) r_j = N_{f(j)}(t)$.

To account for growth in either the number of articles or citations in each field, we assign a growth rate λ_{Ni} to the number of papers published in field i , λ_{ci} to the average number of references per paper in this field, and λ_j to the growth of the focal journal (see Box 1 for definitions of growth). We then get:

$$CI_j = r_j \sum_{i=1}^F \left(\frac{c_i(0)}{\sum_{t=1}^T \lambda_j^{-t}} \cdot \frac{N_i(0) \sum_{t=1}^T (\lambda_j \lambda_{Ni} \lambda_{ci})^{-t} R_{i,r(f(j))}(t)}{\sum_{k=1}^F \left(N_k(0) \sum_{t=0}^{\infty} \lambda_{Ni}^{-t} R_{ik}(t) \right)} \right) \tag{3}$$

From this it can be seen that growing numbers of citations per paper always increases impacts, but the effect of growth of journals or whole fields can vary. A growing subject may increase its impact factor since there are more current references citing a smaller pool of past papers (effect of λ in the denominator), but if citations accumulate slowly, strong growth also means that most papers have appeared too recently to be cited (effect of λ in the numerator). Because of these opposing effects, the total effect of growth rates is likely to be limited, compared with the effect of the total number of citations per paper and the proportion of those falling within the impact factor time window.

In Eqn 3, the impact factor of a given journal is proportional to its scientific relevance r_j , but the value that is reached depends on citation patterns within and among fields. If there are no interdisciplinary references, Eqn 3 reduces to:

$$CI_j = r_j \frac{c_{r(f(j))}(0) \sum_{t=1}^T (\lambda_j \lambda_{Nj} \lambda_{cj})^{-t} R_{f(j),r(f(j))}(t)}{\sum_{t=1}^T \lambda_j^{-t} \sum_{i=0}^{\infty} \lambda_{Ni}^{-t} R_{f(i),r(f(j))}(t)} \tag{4}$$

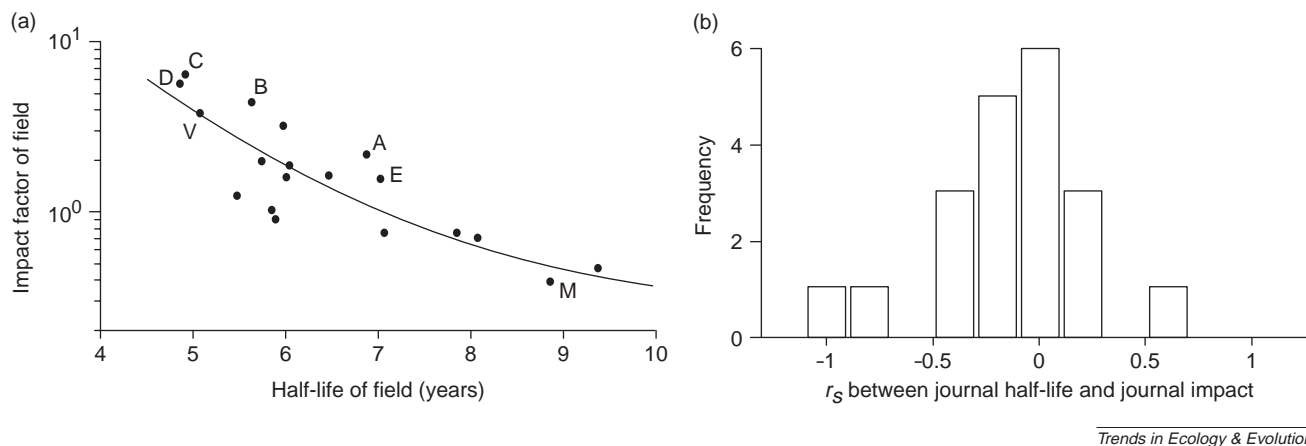
The parameter r_j , which is nearest the idea of ‘quality’ of research, is automatically limited by the scaling $\sum_j n_j(t) r_j = N_{f(j)}(t)$. In other words, the expected r_j of a randomly chosen paper must equal one within every field, and increasing the relevance r_j of all results within the field by the same factor will have no effect at all on the overall impact. Instead, if growth rates are small (each $\lambda \approx 1$), the impact factor of a field is completely determined by the field-specific average of the number of citations in each paper, $c_{r(i)}$, scaled by the fraction of cited papers that were published in the last T years, $\sum_{t=1}^T R(t) / \sum_{t=0}^{\infty} R(t)$. This proves that impact comparisons among fields are meaningless unless the primary interest is to find a complicated way of expressing average number of citable references per paper combined with the proportion of citations that have appeared recently.

quality or number of papers. The same rule must also apply if there are several journals within a subject: the mean citation index for the subject must equal the mean number of citable references per paper within that subject, scaled only by the time span covered by the index. However, the range of impact values can be wider for fields with many journals. Therefore, larger fields are likely to have higher impact factors for their top journals^{2,11}.

Unlike the size of a field, its growth rate has the potential to alter impact factors (Box 2). A high growth rate in the number of papers published annually within a field can increase its impact, but this effect can be reversed if citations tend to accumulate slowly (Box 2). Paradoxically, successful journals often expand, yet this may reduce their impact factor. This is because most papers are cited when they are at the old end of the time

window – having a lower proportion of papers at this end would weight the impact factor unfavourably.

The problem of comparing impacts among fields might be reduced by the fact that comparisons often take place within fields; for example, when comparing the curriculum vitae of applicants for academic positions¹². We agree with this, but we fear that the temptation to use impacts as comparative tools also exists



Trends in Ecology & Evolution

Fig. 1. (a) The impact factor of a field has an inverse relationship to the average half-life of articles within that field ($r_s = -0.91$, $P = 0.009$). The data comprise scientific fields mentioned in the text [astronomy and astrophysics (A), mathematics (M), and five biological fields – developmental biology (D), cell biology (C), virology (V), biochemistry and molecular biology (B) and ecology (E)], in addition to 12 other randomly chosen fields from Journal Citation Reports⁷. Journal half-lives above ten years were truncated to ten years because of the recording practice used; the true half-life for the rightmost fields should thus be greater. After correcting for differences in half-life by the shown 2nd order polynomial fit, the size of a field does not explain variation in impacts: $r_s = 0.09$ ($P > 0.05$) between impact residuals and the number of published articles in a field (even for uncorrected data, $r_s = 0.20$ between impact and size, $P > 0.05$). (b) Frequency distribution of correlations between half-life and impact factor of the same fields. A previous study⁶ showed a correlation between impact factor and turnover, which is inversely related to half-life, in 28 biological and biomedical journals. However, a negative relationship is not consistently found within fields: Spearman correlations between journal half-life and journal impact range from -1 (developmental psychology) to 0.54 (medical laboratory technology).

among fields that appear related, such as the different biological sciences in Fig. 1. Even within a field, there will be variation in the traditions of different subfields, again making comparisons difficult.

For example, within ecology and evolution, there are molecular ecology journals with a typically rapid turnover, and taxonomy journals that have much longer half-lives. Scientists publishing in journals of a subfield that get listed along with another subfield with a shorter half-life or a longer mean number of citable references will suffer from this bias^{1,4,11}. Because long half-lives mean that a large number of citations are not indicated by the impact factor, such biases could be reduced if half-lives were considered an equally important measure of quality as impact factors¹³. However, further complications are found when among-field citations are possible. Citing papers from other fields reduces the impact of the quoting field while increasing the impact of the quoted field (Eqn 3 in Box 2). It is also clear that the impact factor of interdisciplinary journals gets exaggerated in fields that have lower impacts, such as ecology, but is understated for those which otherwise have higher impacts, such as developmental biology.

Comparisons using impact factors have serious implications for ecologists and evolutionary biologists. The constraints of the slow life cycles of many species can make rapid studies impossible, and often details of the species or study site demand reference to older or obscure studies. Improvements such as

relative rankings or impact divided by mean impact for that field are unlikely to offer completely satisfactory solutions simply because fields are never totally isolated in their citation patterns. It seems that any proper comparison must be constrained to journals within a field, and, even then, more fine-scaled divisions are likely to be problematic: the same issues apply at a smaller scale because any field can be subdivided.

Similar problems arise from the practice of counting citations to an individual researcher's work. Furthermore, because of time lags between publication and citation, this tends to underestimate young scientists' work and might overestimate the contribution of individuals with declining productivity. A further concern is that the demand for high impact ratings will subvert science; for example, by placing increasing emphasis on areas where it is possible to complete and carry out research very quickly. This is of considerable concern for ecologists and evolutionary biologists where long-term studies are a fundamental part of the science.

Acknowledgements

We thank Neil Metcalfe and an anonymous referee for very helpful comments, and the TMR programme of the European Commission for funding.

Hanna Kokko

Dept of Zoology, University of Cambridge, Downing Street, Cambridge, UK CB2 3EJ (h.kokko@zoo.cam.ac.uk)

William J. Sutherland

School of Biological Sciences, University of East Anglia, Norwich, UK NR4 7TJ (w.sutherland@uea.ac.uk)

References

- Lowy, C. (1997) **Impact factor limits funding**, *Lancet* 350, 1035
- Seglen, P.O. (1997) **Why the impact factor of journals should not be used for evaluating research**, *Br. Med. J.* 314, 498–502
- Schoonbaert, D. and Roelants, G. (1998) **Impact takes precedence over interest**, *Nature* 391, 222
- Statzner, B., Resh, V.H. and Kobzina, N.G. (1995) **Scale effects on impact factors of scientific journals: ecology compared to other fields**, *Oikos* 72, 440–443
- Peters, R.H. (1991) *A Critique for Ecology*, Cambridge University Press
- Metcalfe, N.B. (1995) **Journal impact factors**, *Nature* 376, 720
- Anon. (1996) *Journal Citation Reports* (Science Edition), Institute for Scientific Information, Philadelphia
- Wiles, A. (1995) **Modular elliptic curves and Fermat's last theorem**, *Ann. Math.* 141, 443–551
- Singh, S. (1997) *Fermat's Last Theorem*, 4th Estate
- Gomperts, M.C. (1968) **The law of constant citation for scientific literature**, *J. Document.* 24, 113–117
- Statzner, B., Resh, V.H. and Kobzina, N.G. (1995) **Low impact factors of ecology journals – don't worry**, *Trends Ecol. Evol.* 10, 220
- Motta, G. (1995) **Journal impact factors**, *Nature* 376, 720
- Linardi, P.M., Coelho, P.M.Z. and Costa, H.M.A. (1996) **The 'impact factor' as a criterion for the quality of scientific production is a relative, not absolute, measure**, *Braz. J. Med. Biol. Res.* 29, 555–561