

Research article

Open Access

The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor

Andy R Weale*^{1,2}, Mick Bailey¹ and Paul A Lear²

Address: ¹Molecular and Cellular Biology, Division of Veterinary Pathology Infection & Immunity, University of Bristol, Langford House, Langford BS40 5DU, UK and ²Department of Surgery, Southmead Hospital, North Bristol NHS Hospitals Trust, Westbury on Trym, Bristol BS10 5NB, UK

Email: Andy R Weale* - andy@weale.org.uk; Mick Bailey - mick.bailey@bristol.ac.uk; Paul A Lear - paul.lear@north-bristol.swest.nhs.uk

* Corresponding author

Published: 28 May 2004

Received: 10 February 2004

BMC Medical Research Methodology 2004, **4**:14

Accepted: 28 May 2004

This article is available from: <http://www.biomedcentral.com/1471-2288/4/14>

© 2004 Weale et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Current methods of measuring the quality of journals assume that citations of articles within journals are normally distributed. Furthermore using journal impact factors to measure the quality of individual articles is flawed if citations are not uniformly spread between articles. The aim of this study was to assess the distribution of citations to articles and use the level of non-citation of articles within a journal as a measure of quality. This ranking method is compared with the impact factor, as calculated by ISI®.

Methods: Total citations gained by October 2003, for every original article and review published in current immunology (13125 articles; 105 journals) and surgical (17083 articles; 120 journals) fields during 2001 were collected using ISI® Web of Science.

Results: The distribution of citation of articles within an individual journal is mainly non-parametric throughout the literature. One sixth (16.7%; IQR 13.6–19.2) of articles in a journal accrue half the total number of citations to that journal. There was a broader distribution of citation to articles in higher impact journals and in the field of immunology compared to surgery. 23.7% (IQR 14.6–42.4) of articles had not yet been cited. Levels of non-citation varied between journals and subject fields. There was a significant negative correlation between the proportion of articles never cited and a journal's impact factor for both immunology ($\rho = -0.854$) and surgery journals ($\rho = -0.924$).

Conclusion: Ranking journals by impact factor and non-citation produces similar results. Using a non-citation rate is advantageous as it creates a clear distinction between how citation analysis is used to determine the quality of a journal (low level of non-citation) and an individual article (citation counting). Non-citation levels should therefore be made available for all journals.

Background

Impact factors have been used to evaluate the quality of journals for decades[1], and they are finding an increasingly influential role within science[2]. Authors and academic institutions are now frequently judged and funded simply on the basis of publications in a high impact journals[2]. Yet as a quality indicator of individuals and insti-

tutions impact factor is frequently criticised and is certainly flawed[3].

The impact factor of a journal is arrived at by a simple calculation. For any particular year (e.g. 2002), it is defined as the number of citations in that year (i.e. 2002), to articles published in the two previous years (i.e. 2001 and

2000), divided by the total number of source articles published in that time. ISI® states that this represents the "frequency with which the 'average article' in a journal has been cited in a particular year". However, citations may not be evenly distributed amongst articles in a journal and a small number of articles probably attract the bulk of citations[3]. If this is true for the majority of journals, the calculation provides a mean number of citations, within a skewed non-parametric population of citations – an intrinsic statistical error. It has also been suggested that half the literature published is redundant as it is never cited[4]. Clearly, suggesting that all articles published in a journal are of similar quality is nonsensical. Even Garfield, the originator of the impact factor, states that it is incorrect to judge an article by the impact factor of the journal[5].

We hypothesised that journals which are of high quality are likely to have few articles that are never cited and measures of non-citation in individual journals may therefore provide an alternative and perhaps more appropriate way of comparing journals. Journals with low levels of non-citation of articles may, incidentally, also be those which attract the most citations. The aim of this study was to assess the distribution of citations and particularly the level of non-citation within 2 areas; one of which contains a majority of basic science articles (immunology) and the other a majority of clinical science articles (surgery). Levels of non-citation are compared with impact factor as a method of measuring quality of journals.

Methods

Journal selection

All journals listed in the 2002 ISI® Journal Citation Report under the subject heading of "Immunology" and "Surgery" were included for analysis. Journals were excluded if the journal had not been published in 2001 or 2000, or were no longer published. For the purpose of this study, we defined journals containing less than twenty percent review articles as being a primary research journal.

Impact factor and citation counts

The ISI® Journal Citation Report of 2002 was interrogated to obtain each journal's impact factor (2002), number of citations (2002) and number of source articles (2000 & 2001).

In October 2003, the number of citations to every article (or up to the maximum obtainable of 500) classified as an original article or review, published in 2001, for each journal was obtained using the ISI Web of Science®. The type of article (review or original research) and its length in terms of number of pages were also retrieved.

The distribution of citations of articles within each journal; the influence of the article type and length on citation, and the relationship between proportion non-cited articles within a journal and its Impact factor were investigated.

We wish to highlight that our data includes citations only to original research and review articles published in 2001 made from the date of publication up to October 2003. The 2002 journal impact factor relates to the number of citations made in 2002 to any type of article published in the two previous years (2000 & 2001). Extraction of year specific citations from the ISI reports is not possible on the scale of this study. Whilst we recognize that this disparity creates limitations to the study, the aim was to examine the hypothesis that a measure of non-citation (of a set articles at any defined point) provides an alternative method of ranking journals. As such we believe it is acceptable to compare of non-citation of the 2001 literature with the ranking produced by the 2002 impact factor.

Statistical Analysis

A one sample Kolmogorov-Smirnov test was used to assess the normality of the distribution of citations within each journal. The Mann-Whitney U Test was used to compare number of pages, citations, and cites/page between groups. Chi squared and Fisher's exact tests were used for tests of proportions, as appropriate. The comparison of ranking of journals produced by impact factor and non-citation was performed using Spearman rank correlation.

Statistical analysis was performed with the Software Package for the Social Sciences (version 10.1 for Windows, SPSS Inc, Chicago, Ill, USA). A p value of <0.05 was considered statistically significant.

Results

The distribution of citation of articles is mainly non-parametric throughout the literature

13125 immunology journal articles published across the 105 journals and 17083 surgical articles in 120 journals were evaluated. The median impact factor of the two groups of journals are 1.895 (IQR 1.283–3.022) for all the immunology journals and 0.881 (IQR 0.568–1.724) for the surgical group ($p < 0.0001$ Mann Whitney). Of these journals, 13 in the immunology group and 5 in the surgical group contained more than twenty percent review articles.

Of all the immunology journals only 18 (17.1%) had a normal distribution of citations to articles, whilst only 9 (7.5%) of the distributions of citations to surgical articles were parametric ($p = 0.026$ χ^2 test). These journals with a normal distribution of citations publish only a small number of articles per year, and account for less than 5%

of the 2001 immunology literature, and less than 2% of the surgical literature. These journals were more likely to contain a high proportion of reviews than be primary research journals ($p < 0.0001$ Fisher's exact).

In the 198 journals which publish the bulk of articles (>95%) the distribution of citations amongst articles was non parametric. Histograms for a sample of 5 primary research immunology journals are shown in Figure 1.

To evaluate whether the distribution of articles was different between journals, the proportion of articles which account for 50% of all of a journal's citations were calculated. In the immunology group, a median of 18% (IQR 15–21) of a journal's articles accounted for 50% of all citations to that journal. A significantly smaller number of articles (median 15%, IQR 13–18) gained over half a journal's citations in the surgical literature (Mann whitney $p < 0.0001$). Figure 2 shows that this figure varied considerably between journals. However there was a significant correlation between impact factor and the proportion of journal articles accounting for the bulk of the citations for both surgical and immunology journals. Yet even in the highest ranked primary research immunology journal, Nature Immunology, just 30 of the 132 articles published in 2001 accounted for over half the citations, and 40% of these were reviews.

Level of citation differs with article type and length

It has been suggested that longer articles and review articles receive more citations[3]. Longer articles were indeed more likely to collect citations but this association was weak (Spearman rho = 0.286 and 0.335 for immunology and surgery journals respectively, $p < 0.0001$). Tables 1 and 2 show the number of pages and citations per article, and the number of citations per page for reviews and original articles. This study confirms the commonly held view that review articles attract more citations. Articles in immunology journals receive more citations per page than surgical journals ($p < 0.0001$ Mann Whitney).

Ranking journals by levels of non-citation correlates with impact factor ranking

Of the 30208 articles, 7353 (24.3%) are yet to be cited. The level of non-citation was significantly lower for reviews (14.8%) compared to original articles (24.9%) ($p < 0.0001$ χ^2 test $df = 2$).

The median percentage of un-cited articles in each immunology journal was 17.6% (IQR 10.0–32.0). This was significantly less than the median level of non-citation in each surgical journal (32.8% IQR 21.0–50.0, Mann Whitney $p < 0.0001$). Figure 3 shows that the distribution of percent un-cited articles varied significantly with the

impact factor of the journal for both immunology and surgical journals.

Discussion

Our results demonstrate that the distribution of citations differs between types of article and journals within journal subject areas. It appears that the journals with the highest impact factors generally had a lower proportion of articles which are never cited. This was reproducible between two subject fields which overlap with only two journals (Transplantation and Transplantation Proceedings).

The quality of a journal is difficult to assess objectively and perhaps impossible to define numerically. As such any method is likely to be open to criticism. Quality measures are useful to publishers, advertisers, librarians, editors and authors alike[6]. Editors are often stirred to put pen to paper when a journal's impact factor rises[7,8], whilst authors may use impact factors to decide were to submit scientific research[5]. Whether the increasing ease of access to the abstracts and full text of articles by the internet and electronic publications, particularly open access, will change the importance attached to such measures is unknown[9]. Furthermore simple citation measures may more accurately reflect the usefulness of an article to another authors' work rather than its quality. It has been previously recognised that citation rate and impact factors alter depending on the field of a journal, with basic science journals having higher, impact factors than clinical medicine journals[3]. This is supported by the observation in this study that impact factors were higher and levels of non-citation lower, in the immunology literature compared to the surgical literature. There may also be differences in the relevance of citation counting between clinical and scientific research. It is reasonable to hypothesise that pure clinicians may read articles and journals which influence their clinical practice but never cite this work themselves. Some important clinical papers which guide clinical practice may gain a high number of citations; for example the North American Symptomatic Carotid Endarterectomy Trial[10] has gained over 500 citations and the UK small aneurysm trial[11] over 170. Yet it is easy to find papers which probably do not influence clinical practice as widely, published at similar times in the same journals, which gather large numbers of citations. For example a clinical report of 'Buffalo Hump' in males with HIV infection[12] has accumulated over 280 citations. However it is not known whether research which changes clinical practice is cited more frequently than research which does not. Citation practices of authors may also be influenced by factors other than quality, including language of publication[4] and personal choice[13]. Across the literature, authors are more likely to cite longer articles and review articles.

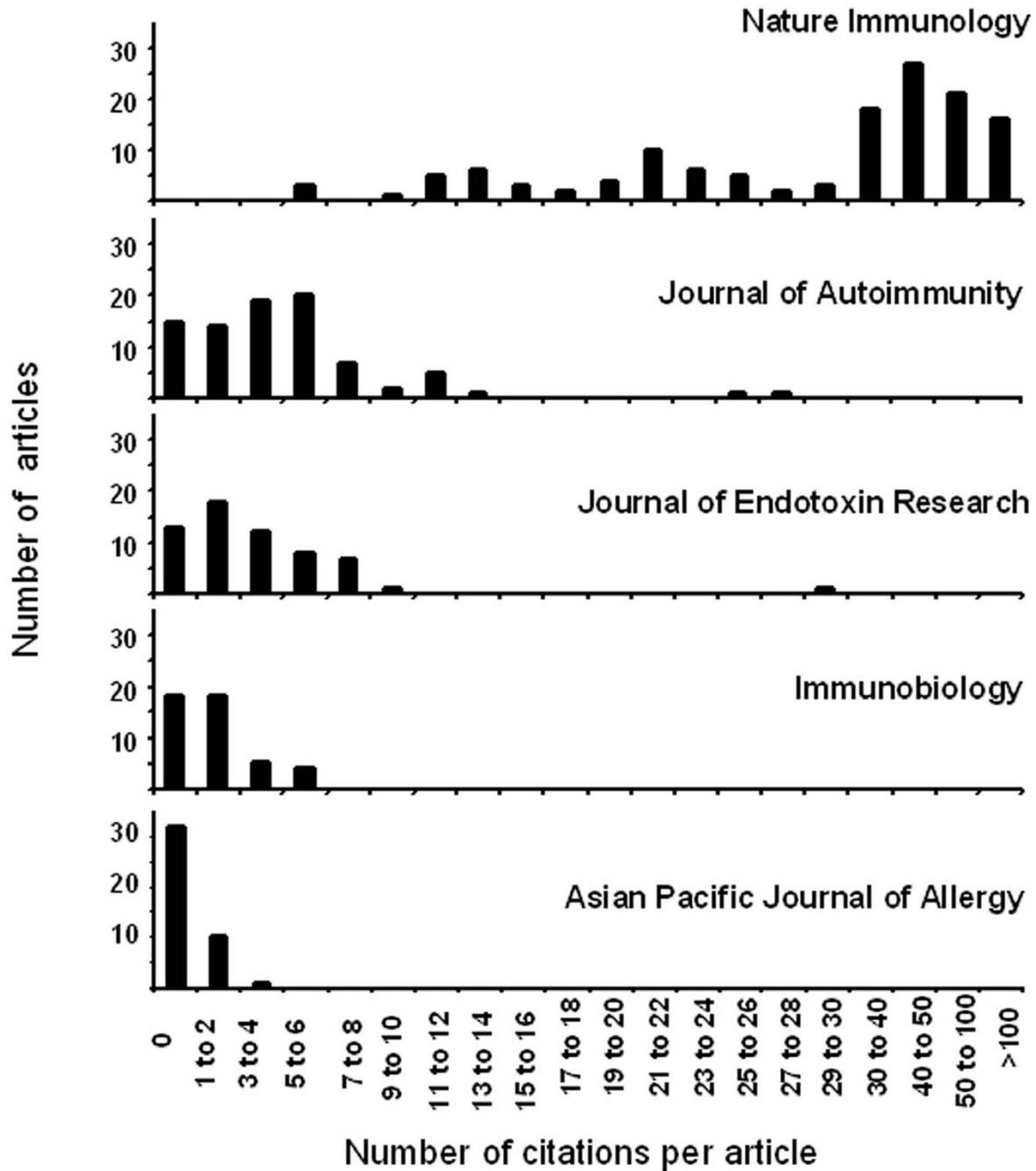


Figure 1
The distribution of citations amongst articles for 5 primary research immunology journals. The journals represent the journals with the highest (Nature Immunology 27.868), lowest (Asian Pacific Journal of Allergy 0.179), median (Journal of Endotoxin Research 1.893), 25th (Journal of Autoimmunity 2.812) and 75th (Immunobiology 1.319) centile impact factors of primary research immunology literature (n = 92).

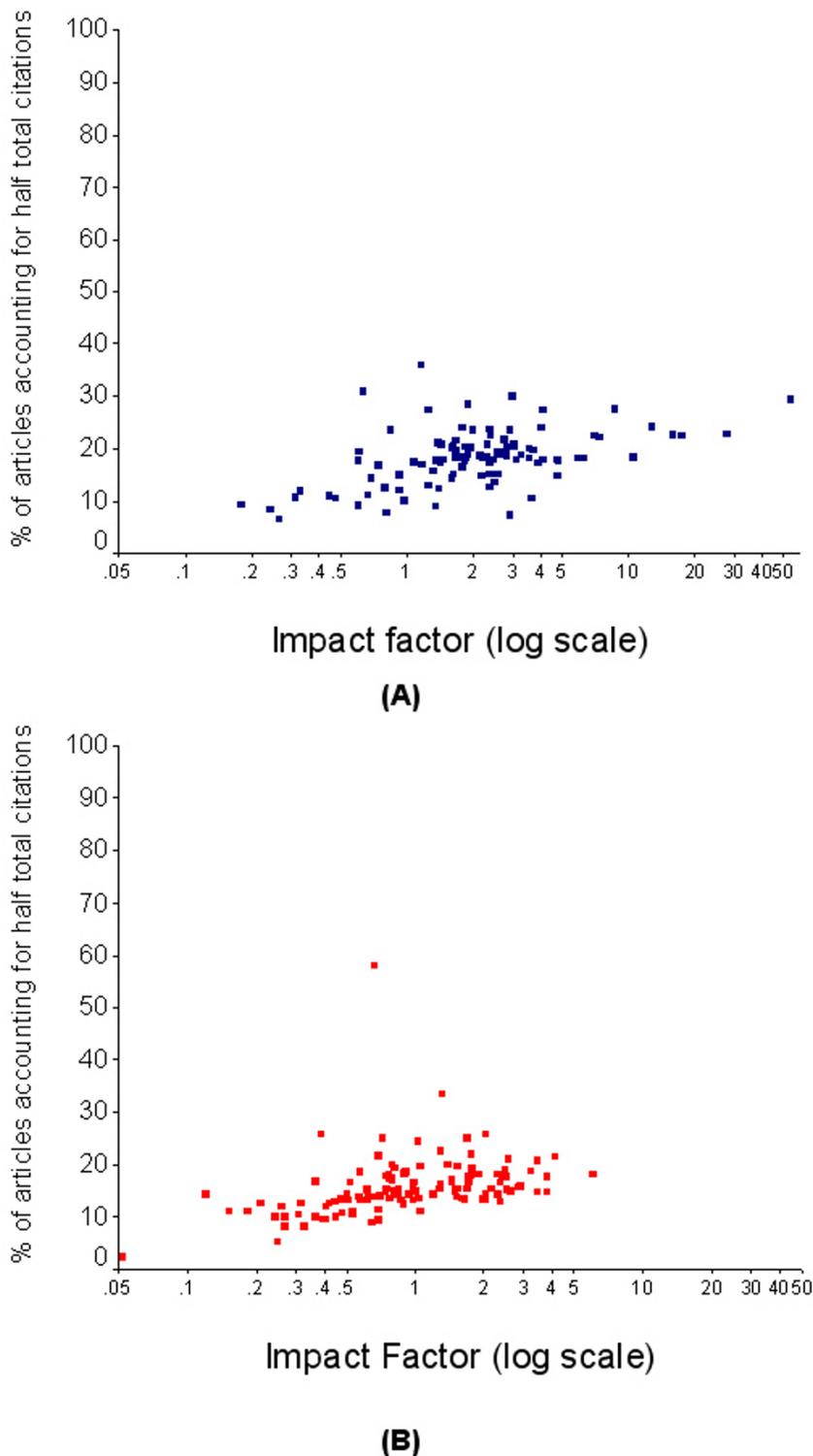


Figure 2
Relationship between impact factor and the distribution of citations to articles in a journal The proportion of articles which accounted for 50% of all journal citations is compared with impact factor (log scale) for immunology (a) and surgical (b) journals. There were significant correlations. For immunology journals (n = 105), Spearman Rank correlation rho = 0.463 p < 0.0001. For surgical journals (n = 120), Spearman Rank correlation rho = 0.556 p < 0.0001.

Table 1: Median numbers of pages, citations and citations per page for all articles in the immunology journals of 2001 (n = 105).

	Original articles	Reviews	p value*
Total	11755	524	
Number of pages per article	7 (5-9)	9 (7-13)	<0.0001
Number of citations per article	3 (1-6)	6 (2-13)	<0.0001
Number of citations per article page	0.43 (0.14-1)	0.6 (0.25-1.31)	<0.0001

*Mann-Whitney U-test

Table 2: Median numbers of pages, citations and citations per page for all articles in the surgical journals of 2001 (n = 120).

	Original articles	Reviews	p value*
Total	16452	631	
Number of pages per article	5 (4-7)	8 (5-11)	p < 0.0001
Number of citations per article	1 (0-4)	2 (1-6)	p < 0.0001
Number of citations per article page	0.286 (0-0.667)	0.333 (0.083-0.707)	p = 0.013

*Mann-Whitney U-test

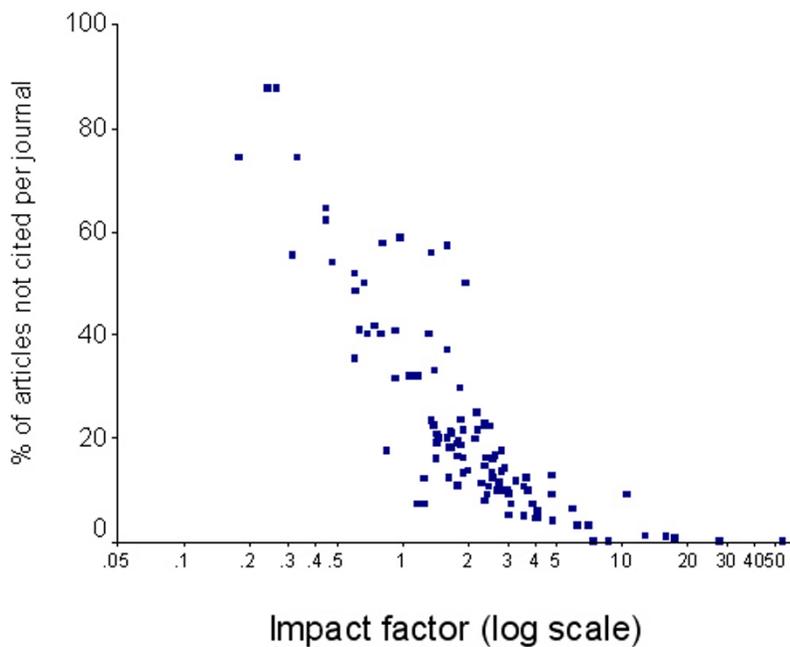
Despite these limitations, citation counts provide a convenient and objective method of ranking articles and journals. It is therefore important to use the most appropriate and transparent way of communicating this information, particularly if such rankings are used to define quality.

The criticism of the impact factor itself has grown as its influence increases. Articles such as editorials, letters and news items are classified as "non-source" items and as such does not count towards the total number of articles used to calculate the impact factor. However, such items may attract numerous citations which are counted towards a journal's impact factor. Journals may increase the number of non-source items to artificially increase impact factors[14]. It is also suggested that the calculation provides a method for comparing journals regardless of their size[5]. However journal size may be a confounding factor- journals publishing more articles tend to have higher impact factors per se[15]. Small journals may be disadvantaged by this bias. Most importantly impact factor does not communicate any information about the citation distribution to the reader.

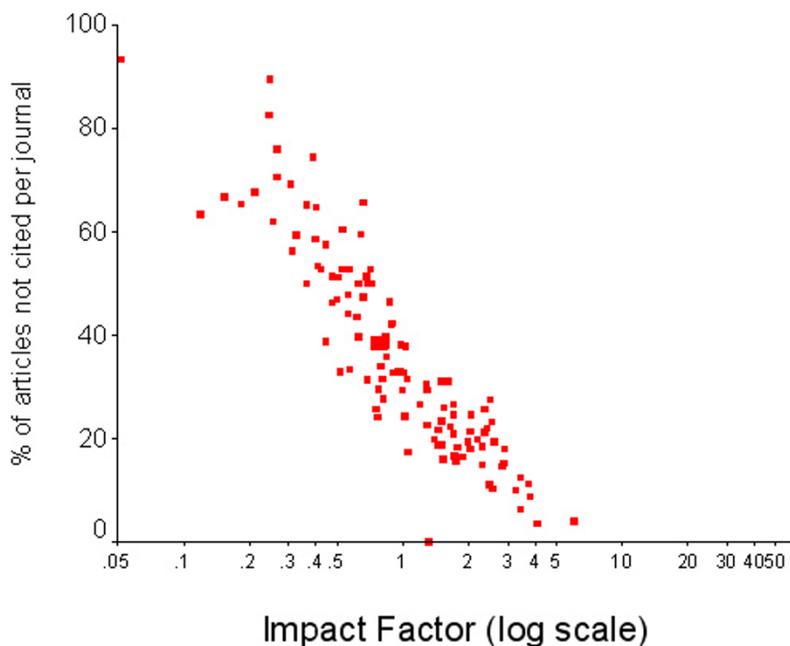
Which is, then, the more appropriate method to measure quality; levels of non-citation or impact factor? Clearly this depends on the definition of quality. One can define a quality journal, in terms of non-citation, as one which maximises the amount of useful, interesting and original information per issue. What is the definition of quality as measured by impact factor? When the calculation is studied it is clear that impact factor represents a mean number

of citations; yet we have demonstrated that the distribution of citations to articles within the vast majority of journals is non-parametric. Statistically, at the very least, the impact factor should represent the median number of citations to articles and not the mean. It is of great concern that the tool which is accepted as a measure of journal quality contains the type of fundamental statistical error which would make most editors and peer reviewers recoil.

The non-parametric distribution of citations to articles lies at the heart of the problem with impact factors. A journal which contains a handful of very useful articles with a large amount of articles which are not cited subsequently may have the same impact factor as one with a small number of citations spread evenly across most of its articles. It was observed that journals with high impact factors do tend to have a broader distribution of citations amongst articles (although rarely Gaussian) and lower levels of non-citation. It would be easy to conclude from this that impact factor therefore also reflects non-citation and it is unnecessary to consider different methods of ranking journals. Using a non-citation rate as a measure of the quality of a journal does have advantages over impact factor, particularly for contributors and institutions. Firstly, the definition of quality is explicit and logical. Most importantly however, it creates a clear distinction between how citation analysis is used to measure the quality of a journal (low level of non-citation) and an individual piece of work (citation counting). This will hopefully remove the temptation to use a journal's ranking to judge individual articles. Articles are, of course, best



(A)



(B)

Figure 3
Relationship between impact factor and the level of non-citation within a journal The proportion of articles which have not been cited is compared with impact factor (log scale) for immunology (a) and surgical (b) journals. There were significant correlations. For immunology journals (n = 105), Spearman Rank correlation $\rho = -0.854$ $p < 0.0001$. For surgical journals (n = 120), Spearman Rank correlation $\rho = -0.924$ $p < 0.0001$.

assessed by reading them, but they may be evaluated by counting citations[5]. Although this is a less than ideal way of measuring quality, it may be preferable to the current method of assuming that an article is good because it is published in a journal which attracts many citations, even though these citations are unevenly dispersed.

In this study we have defined our measure of journal quality in terms of non-citation by evaluating the non-citation of one year's literature (2001) from publication to present day, due to the practicalities of the data retrieval. Should publication of non-citation rates be embraced then this information could be presented in a number of ways. Firstly the level of non-citation within the current year to the previous two years articles could be presented alongside the impact factor. However this does not overcome the problems of temporal bias produced by only reporting the citation statistics relating to 2 recent years[3]. The level of non-citation could therefore also be reported yearly or even continuously for every previous year of each journal. This would provide an index which takes into account every citation made to a journal rather than just those made in a short period of time following publication.

Impact factor does have some advantages over non-citation measures for a handful of journals. For those journals which have no un-cited literature (2.2% of journals included in this report) the impact factor offers a further way of discriminating between journals, yet interestingly of the 5 journals with no un-cited literature only one, Nature Immunology, was a primary research journal. Perhaps non-citation is most useful for ranking the primary research literature. As reviews and original articles attract different levels of citation, it may be most appropriate to use the level of non-citation of original articles as the measure of a journal's quality. This would also mean that the citation of non-source items and publication of numerous reviews would not improve a journal's ranking as it may do using impact factor. This is an area for debate. Furthermore given that non-citation and citation practices are different in individual subject fields the measure of non-citation is probably no more valid than impact factor for comparing journals between fields. However, within individual subject fields, non-citation provides a more logical and explicit measure of a journal's quality than impact factor.

Conclusions

Ranking journals by impact factor and non-citation does produce similar results. Data about un-cited literature is currently difficult to obtain, even though it may represent a more relevant and logical measure of the quality or usefulness of a journal than the impact factor. Non-citation levels should therefore be made available for all journals. Accepting the importance and incorporating such infor-

mation about un-cited literature into reports will protect the isolated impact factor from more criticism and provide a clearer measure of quality.

Competing interests

None declared.

Authors' contributions

ARW suggested the initial hypothesis, designed the study, acquired and analysed all data presented. The first drafting of the manuscript was undertaken by ARW. MB and PAL were equally involved in helping to develop the hypothesis and the critical appraisal/re-writing of the manuscript. All authors have read and approved the final manuscript. ARW takes overall responsibility for the integrity of the work.

References

1. Garfield E: **Citation analysis as a tool in journal evaluation.** *Science* 1972, **178**:471-479.
2. Adam D: **The counting house.** *Nature* 2002, **415**:726-729.
3. Seglen PO: **Why the impact factor of journals should not be used for evaluating research.** *BMJ* 1997, **314**:498-502.
4. Granda Orive JI: **Reflections on the impact factor.** *Arch Bronconeumol* 2003, **39**:409-417.
5. Garfield E: **Journal impact factor: a brief review.** *CMAJ* 1999, **161**:979-980.
6. Lundberg G: **The "omnipotent" Science Citation Index impact factor.** *Med J Aust* 2003, **178**:253-254.
7. **Summer Dreams.** *Nat Immunol* 2003, **4**:715.
8. Murie J: **Changing times for BJS.** *Br J Surg* 2003, **90**:1-2.
9. Brunstein J: **End of impact factors?** *Nature* 2000, **403**:478-478.
10. Barnett HJ, Taylor DV, Eliasziw M, Fox AJ, Ferguson GG, Haynes RB, Rankin RN, Clagett GP, Hachinski VC, Sackett DL, Thorpe KE, Meldrum HE, Spence JD: **Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators.** *N Engl J Med* 1998, **339**:1415-1425.
11. **Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. The UK Small Aneurysm Trial Participants.** *Lancet* 1998, **352**:1649-1655.
12. Lo JC, Mulligan K, Tai VW, Algren H, Schambelan M: **"Buffalo hump" in men with HIV-1 infection.** *Lancet* 1998, **351**:867-870.
13. Borner K, Maru JT, Goldstone RL: **The simultaneous evolution of author and paper networks.** *Proc Natl Acad Sci U S A* 2004, **101** Suppl 1:5266-5273.
14. Gowrishankar J, Divakar P: **Sprucing up one's impact factor.** *Nature* 1999, **401**:321-322.
15. Weale AR, Lear PA: **Randomised controlled trials and quality of journals.** *Lancet* 2003, **361**:1749-1750.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/4/14/prepub>