



Hidden dangers of a 'citation culture'

Peter A. Todd^{1,2,*}, Richard J. Ladle²

¹Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543, Singapore

²Oxford University Centre for the Environment, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK

ABSTRACT: The influence of the journal impact factor and the effect of a 'citation culture' on science and scientists have been discussed extensively (Lawrence 2007; *Curr Biol* 17:R583–585). Nevertheless, many still believe that the number of citations a paper receives provides some measure of its quality. This belief may be unfounded, however, as there are 2 substantial areas of error that can distort a citation count or any metric based on a citation count. One is the deliberate manipulation of the system by scientists trying to ensure the highest possible number of cites to their papers; this has been examined elsewhere (Lawrence 2003; *Nature* 422:259–261). The second area of inaccuracy is inherent to how papers are cited, indexed and searched for. It is this latter, lesser known, source of error that we will investigate here.

KEY WORDS: Bibliometrics · Citation counts · h-index · g-index · Assessment · Scientists

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

The huge variety of scientific work-practices combined with the difficulty of objectively assessing the quality of a scientist's research output has renewed interest in bibliometric methods of quality assessment. For example, the UK government is moving to a system of citation-based metrics in place of the current unwieldy panel-based peer-review system for allocating research funding to university departments (Todd & Ladle 2008). The underlying logic of such a move is that the 'value' of a scientific paper will be more effectively captured by the number of citations it accrues than by a panel of experts who may miss the significance or importance of work that is somewhat outside their realm of expertise (Taylor et al. 2008, this Theme Section[TS]). Furthermore, citations seem to be superficially less subject to manipulation and game-playing than, for example, paper length, journal quality, or publication output.

The fundamental assumption of any bibliometric system is that a citation represents a unit of (positive) quality, but it is not clear that the association between the quality of a paper and the number of citations to it is strong enough to form the basis of a rigorous assessment of individuals or institutions. While the deficiencies of using a journal's impact factor as a surrogate metric for the quality of the papers it publishes are well understood (e.g. Garfield 1996, Campbell 2008 [this TS]), citation

counting as a metric of an author's 'contribution' has not been subject to similar levels of scrutiny.

It is well known that authors will go to great lengths to publish their work in high impact journals (e.g. Lawrence 2003), and similar strategies could be employed to help ensure their papers get cited (e.g. 'citation-bartering', Lawrence 2007). Any such manipulation of the system clearly raises doubts about the appropriateness of citation-based assessment. Here, however, we will focus on the intrinsic weaknesses of counting citations as a measure of quality and argue that these alone are enough to cast serious doubt on the validity of all citation-based bibliometrics. Of course, it should be kept in mind that scientists are under increasing pressure and time constraints (Ladle et al. 2007) and, while this does not excuse poor referencing practices, it does go some way to explaining the prevalence of such practices and the difficulties that might be involved in improving the present situation.

MISCITATION/MISQUOTATION

Simkin & Roychowdhury (2003) claim that only ~20% of papers are actually read by the authors citing them, based on their study of how misprinted citations proliferate through the physics and engineering literature. As copying citations from other sources does not

*Email: dbspat@nus.edu.sg

automatically correlate with whether those papers have in fact been read, their figure is probably an over-estimation. Where it has been studied in more detail, the use of inappropriate citations to support assertions in peer-reviewed articles has been found to be widespread. For example, the number of miscitations (also called misquotations) in the biomedical literature varies between 6% in radiology journals (Hansen & McIntire 1994) and 35.2% in emergency medicine journals (Goldberg et al. 1993).

As no such information existed for ecological sciences (or most other disciplines) we (Todd et al. 2007) examined 306 papers from 51 ecology journals with an impact factor >1. From each paper, we randomly selected 1 citation from the reference list and identified the assertion it was meant to support; we then read the cited article to determine its appropriateness. Overall, we found that only 76.1% of citations clearly supported the assertion they were intended to reinforce. Of the remainder, the support was equivocal (11.1%) or absent (7.2%). The remaining 5.6% of the cases were classified as 'empty', i.e. not a reference to primary research but to a secondary source such as a review, or an article's Introduction (Harzing 2002).

It is difficult to identify what is driving such high rates of miscitation, although it is clear that it is endemic. It would be interesting to know whether there is a trend towards a greater or lesser degree of miscitation with the increasing availability of electronic literature. Even though it is now trivially easy to identify countless references for almost any subject area, many universities have limited subscriptions to electronic journals — thus limiting the access scientists have to a few sentences in an abstract. These can become tantalizing, especially if the full paper cannot be sourced from the author's website or through direct contact. The temptation to cite such incomplete sources, however, should be resisted.

What is the influence of miscitation on bibliometrics? Papers are counted when they should not be (because they were not relevant), whereas the research that should have been cited is not duly credited.

CITING PAPERS THAT ARE 'WRONG'

Most authors will at some time refer to research that contradicts their own. In the majority of instances the cited work will represent good science, simply with different results, but there will also be cases where the 'other' study is just plain wrong! Of course, papers cited to support an assertion might also be erroneous. Here, the author has usually failed to see the flaw, although there are almost certainly instances where an unscrupulous author may deliberately cite a supportive paper even though he or she doubts the findings.

How do papers that are wrong still get cited in good faith? A good example is that of 'self-perpetuating myths and Chinese whispers' (Harzing 2002, p. 144) where 'facts' have become adopted into the literature even though they have deviated significantly from the original publication. Instances can be found in many fields and include such diverse examples as the distortion of expatriate failure rates (Harzing 2002), ant extinctions in Madeira (Wetterer 2006) and details of a hand surgery study (Porrino et al. 2008).

The increasing use of meta-analysis may also unwittingly be promoting this type of error because, unlike narrative reviews, meta-analyses do not generally distinguish between 'good' and 'bad' papers. Furthermore, some meta-analyses rely on secondary data and may therefore unintentionally misrepresent research findings. For example, Whittaker & Heegaard (2003) re-analysed 8 case studies from a highly cited meta-analysis of the productivity–diversity relationship (Mittelbach et al. 2001) published in the well-regarded journal *Ecology*. They found that the original authors had failed to classify the correct statistical shape of the relationship in every single one of these studies. Interestingly, this discovery does not seem to have stopped the paper attracting large numbers of citations.

Finally, it is possible for a paper to be cited because it is an example of how not to do something, or how mistaken someone can be. Garfield (1979, p. 244) claims that such negative citations are 'more theoretical than real'. Underwood (2004, p. 284), on the other hand, states: 'longevity may be conferred on a publication because it is so truly bad that people keep on finding new things wrong with it!' In such a case, high citations would indicate the exact opposite of what they are supposed to be measuring.

SEARCHING AND ACCESSING

Not all journals are equally accessible and it seems reasonable to assume that papers that are hard to obtain will be cited less. This does not reflect scholarship, but rather the decisions of libraries (whose choices are usually limited by budget constraints) or whether an author can afford open access. In libraries, priority may be given to a core of high impact journals (Taylor et al. 2008) so, indirectly, the impact factor of a journal (which is generally considered a poor way to evaluate a scientist's performance) is affecting the likelihood of an individual's paper being cited. Furthermore, Thomson ISI's Web of Science (www.isiwebknowledge.com/), commonly used for calculating the numbers of citations for papers, only counts those from papers within the journals that it lists. For example, Stergiou & Tsikliras (2006, p. 16) found that 'searching

only ISI's database produces an under-representation of the scientific output of professionals studying marine ecosystems.' In other words, if your paper is not published in certain select journals, it does not count.

Even if a journal is available, the papers it contains can be overlooked. When researching a minor aspect of a study, an author might not conduct the most thorough search possible. Under these circumstances, how easily and/or frequently a paper is encountered may influence its chances of being cited more than how appropriate it is. A paper that appears promptly in search engines could simply have an effective title or key words; a more suitable article can be missed because it includes terms that are searched for less often.

The influence of coverage by the mass media on scientific citations is almost impossible to quantify but circumstantial evidence suggests that it most likely has some effect. For example, a paper by Thomas et al. (2004) on extinction risk and climate change generated worldwide media interest primarily due to a poorly phrased press release and some hyperbolic reporting (Ladle et al. 2004, 2005). The paper has gone on to become a citation classic despite various conceptual and methodological weaknesses (Thuiller et al. 2004).

COLLECTION AND CALCULATION

There is some uncertainty as to exactly how citation numbers are identified by entities such as Thomson ISI's Web of Science when determining the impact factor of journals (e.g. Rossner et al. 2007). Referring to *Nature's* efforts to independently work out their own impact factor, Campbell (2008, p.6) observes 'the numbers quoted in calculating the impact factor [by ISI] are highly questionable'. This issue of 'questionable numbers' will also apply to institutes that rely on similar sources for tallying citations when appraising a member of academic staff. The number of citations a paper receives also depends on which electronic system is used; for example, Meho & Yang (2007) noted major differences among Google Scholar, Web of Science and Scopus. When evaluating Google Scholar (GS) as a citation analysis tool, Harzing & van der Wal (2008, this TS, p. 62) note,

'More generally, citations are subject to many forms of error, from typographical errors in the source paper, to errors in GS parsing of the reference, to errors due to some non-standard reference formats. Publications such as books or conference proceedings are treated inconsistently, both in the literature and in GS. Thus, citations to these works can be complete, completely missing, or anywhere in between.'

We think it is likely these issues will apply to all citation search engines and together represent an undesirable degree of uncertainty.

ERRORS IN REFERENCE LISTS

The 'typographical errors in the source paper' mentioned by Harzing and van der Wal (2008) in the previous section present a relatively well-studied phenomenon. There are a number of authors who have published papers detailing typographic and other such mistakes — often called 'citation errors' — found in the reference lists of journals. Taking data from 5 separate biomedical journal studies (DeLacey et al. 1985, Hansen & McIntire 1994, Fenton et al. 2000, Gosling et al. 2004, Lukiç et al. 2004), the mean number of incorrectly referenced citations is an impressive 34.28%.

Such errors can make retrieval of the desired article difficult and sometimes impossible, thus lowering the chance of the incorrectly referenced paper being used and cited. More importantly, if a scientist's contribution is to be assessed by counting citations, they must be retrievable. A citation that does not show up on a database because the author's name is misspelled, for instance, will be overlooked. It would be relatively straightforward to correct these types of error, but the time-consuming nature of the process means that this probably will not happen until there is greater incentive for publishers to induce change or software is developed that automatically checks references for mistakes. Ultimately, the responsibility falls on authors to make sure that their references are correct, as made clear in the International Committee of Medical Journal Editors' guidelines: 'The references must be verified by the author(s) against the original documents' (International Committee of Medical Journal Editors 1999, p. 70).

CONCLUSIONS

The errors and inconsistencies we list in this article are by no means mutually exclusive. They can lead to miscalculation and inaccuracies that can either increase or decrease the actual number of citations a paper receives. Of course, one could contend that these errors are across the board and therefore immaterial, but the 'inaccuracy of the method was the same across all samples' is very rarely an acceptable argument in science and there is no good reason to suggest it should be applied to bibliometrics. Furthermore, it is unlikely that these errors really do manifest equally. For instance, Kotiaho et al. (1999) detail how names from unfamiliar languages are often inputted incorrectly (and are therefore less likely to be picked up in citation analyses), thus introducing geographical bias against non-English speaking countries.

Counting citations, or use of indices based upon such counts (e.g. h-index, g-index), superficially appears to be an objective and efficacious method of quantifying

the 'quality' of a paper, and thus its author/s. Unfortunately, citation counts can also reflect a whole suite of errors and artifacts that distort the metric to the point that it should not be taken at face value. Importantly, all the problems we have discussed are inherent in current normative citing (and citation counting) practices—anything that authors try to do in order to manipulate the system (e.g. Lawrence 2007) will compound the errors listed here. The combination of intrinsic (mostly accidental) error plus deliberate error in the form of author 'game-playing' is all transferred to whatever metric is adopted by the university or institution or grant awarding body that is determining the future prospects of the individual scientist. This is a cause for serious concern. Our paper supports Lawrence's (2007, p. R583) view that impact factors and citations are 'dodgy evaluation criteria', and we strongly advise against a system that wholly relies upon them to evaluate a scientist's contribution. Nevertheless, we acknowledge that there are also deficiencies in review panels and understand that the demand for quantitative measures will undoubtedly remain high. Given the complexity of the assessment task, a multifaceted process that incorporates bibliometrics and peer-appraisal as well as other indicators (e.g. grant capture), similar to that advocated by Butler (2008, this TS), is a compromise that might be palatable to all parties involved.

LITERATURE CITED

- Butler L (2008) Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework. *Ethics Sci Environ Polit* 8: 83–92
- Campbell P (2008) Escape from the impact factor. *Ethics Sci Environ Polit* 8:5–7
- DeLacey G, Record C, Wade J (1985) How accurate are quotations and references in medical journals? *Br Med J (Clin Res Ed)* 291:884–886
- Fenton JE, Brazier H, De Souza A, Hughe JP, McShane DP (2000) The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clin Otolaryngol* 25:40–44
- Garfield E (1979) *Citation indexing: its theory and application in science, technology, and humanities*. John Wiley & Sons, New York
- Garfield E (1996) How can impact factors be improved? *Br Med J* 313:411–413
- Goldberg R, Newton E, Cameron J, Jacobson R, Chan L, Bukata WR, Rakab A (1993) Reference accuracy in the emergency medicine literature. *Ann Emerg Med* 22: 1450–1454
- Gosling CMcR, Cameron M, Gibbons PF (2004) Referencing and quotation accuracy in four manual therapy journals. *Man Ther* 9:36–40
- Hansen ME, McIntire DD (1994) Reference citation in radiology: accuracy and appropriateness of use in two major journals. *Am J Roentgenol* 163:720–723
- Harzing A (2002) Are our referencing errors undermining our scholarship and credibility? The case of expatriate failure rates. *J Organ Behav* 23:127–148
- Harzing A, van der Wal R (2008) Google Scholar as a new source for citation analysis. *Ethics Sci Environ Polit* 8: 61–73
- International Committee of Medical Journal Editors (1999) Uniform requirements for manuscripts submitted to biomedical journals. *Med Educ* 33:66–78
- Kotiaho JS, Tomkins JL, Simmons LW (1999) Unfamiliar citations breed mistakes. *Nature* 400:307
- Ladle RJ, Jepson P, Araujo M, Whittaker RJ (2004) Dangers of crying wolf over risk of extinctions. *Nature* 428:799
- Ladle RJ, Jepson P, Whittaker RJ (2005) Scientists and the media: the struggle for legitimacy in climate change and conservation science. *Interdiscip Sci Rev* 30: 231–240
- Ladle RJ, Malhado ACM, Todd PA (2007) Come all ye scientists, busy and exhausted. O come ye, O come ye, out of the lab. *Nature* 450:1156
- Lawrence PA (2003) The politics of publication. *Nature* 422: 259–261
- Lawrence PA (2007) The mismeasurement of science. *Curr Biol* 17:R583–R585
- Lukić IV, Lukić A, Glunaić V, Katavić V, Vuāenik V, Marušić A (2004) Citation and quotation accuracy in three anatomy journals. *Clin Anat* 17:534–539
- Meho LI, Yang K (2007) Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *J Am Soc Inf Sci Technol* 58:2105–2125
- Mittelbach GG, Steiner CF, Scheiner SM, Gross KL and others (2001) What is the observed relationship between species richness and productivity? *Ecology* 82:2381–2396
- Porrino JA Jr, Tan V, Daluiski A (2008) Misquotation of a commonly referenced hand surgery study. *J Hand Surg* 33:2.e1–2.e9
- Rossner M, Van Epps H, Hill E (2007) Show me the data. *J Cell Biol* 179:1091–1092
- Simkin MV, Roychowdhury VP (2003) Read before you cite! *Complex Syst* 14:269–274
- Stergiou KI, Tsikliras AC (2006) Underrepresentation of regional ecological research output by bibliometric indices. *Ethics Sci Environ Polit* 2006:15–17. Available online at: www.int-res.com/articles/esepp/2006/E68.pdf
- Taylor M, Perakakis P, Varvara T (2008) The siege of science. *Ethics Sci Environ Polit* 8:17–40
- Thomas CD, Cameron A, Green RE, Bakkenes M and others (2004) Extinction risk from climate change. *Nature* 427: 145–148
- Thuiller W, Araújo MB, Pearson RG, Whittaker RJ, Brotons L, Lavorel S (2004) Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature* 43, doi:10.1038/nature02716
- Todd PA, Ladle RJ (2008) Citations: poor practices by authors reduce their value. *Nature* 451:244
- Todd PA, Yeo DCJ, Li D, Ladle RJ (2007) Citing practices in ecology: can we trust our own words? *Oikos* 116: 1599–1601
- Underwood AJ (2004) It would be better to create and maintain quality rather than worrying about its measurement. *Mar Ecol Prog Ser* 270:283–287
- Wetterer JK (2006) Quotation error, citation copying, and ant extinctions in Madeira. *Scientometrics* 67:351–372
- Whittaker RJ, Heegaard E (2003) What is the observed relationship between species richness and productivity? *Comment. Ecology* 84:3384–3390