OPINION ARTICLE
doi: 10.2306/scienceasia1513-1874.2010.36.001

ScienceAsia 36 (2010): **1–5**

On the current obsession with publication statistics

Michael A. Allen

Physics Department, Faculty of Science, Mahidol University, Rama 6 Road, Bangkok 10400, Thailand

e-mail: frmaa@mahidol.ac.th

ABSTRACT: Crude publication statistics such as publication counts and impact factors are routinely being employed to assess individuals and institutions. Although they can play a role in an approximate preliminary assessment, using them for anything more is inappropriate due to their over-simplicity and ease of manipulation. Furthermore, it is argued that rewarding scientists for achieving high scores in such number-based evaluations ultimately leads to a slowing of scientific progress. Suggestions are given on how reliance on statistics can be reduced and their manipulation discouraged.

KEYWORDS: citations, impact factor, h-index, bibliometrics, research assessment

INTRODUCTION

It is clear that the use of statistics relating to the published output of individuals and institutions is not declining, in spite of the increasing consensus on the glaring inadequacies of such measures. Various university websites list publications along with numbers of citations and a growing array of statistics associated with the journals the articles are published in. Some even display graphs of numbers of publications over time alongside those of rival universities, and some faculties allocate part of the funding for each department in direct proportion to the number of papers they published in the previous year. Both universities and grant-awarding bodies are increasingly offering financial rewards for each paper published and now, presumably to decrease the likelihood of a paper being missed during database searches, some funding agencies insist that to earn the reward, the author must (dishonestly) include the agency in their list of affiliations rather than just mention them in the acknowledgements. In grant applications and final reports one is often required to give the journal impact factors for expected or actual papers published, suggesting that the funding agency takes these values seriously. Finally, in some countries it appears that when assessing research output for the award of positions, promotions, or prizes at the national level, more weight is attached to the number of publications than to the quality of the research done by the candidate.

This article first gives a brief review of why it is unfair to judge individuals, departments, or institutions mainly on various commonly used publication statistics, and some improved measures are mentioned. However, the principal goal here is to look at why an over-reliance on or highlighting of publication statistics (and the cruder statistics in particular) is ultimately damaging to the progress of science, and recommendations are made for countering this.

PUBLICATION STATISTICS

Publication count

One of the simplest and probably the most misused methods of evaluation is simply to count the number of publications. This has several obvious drawbacks. First, it does not take into account multiple authors; one should instead use a count of fractional papers where each paper counts as the reciprocal of the number of authors, or, if the information can be obtained, the proportion of the work done by the author in question. The latter approach, if it can be accomplished, is evidently a lot fairer. One hears of authors being added to papers in return for financial compensation or for political reasons. Suspicions are aroused on seeing a straightforward short paper authored by more than just the student and their perfectly competent supervisor – did the other authors really contribute significantly? In some institutions in Germany it is still unthinkable for the head of a laboratory not to exert their territorial right to be included in the author list, no matter how miniscule their involvement.

Even a fractional count is fundamentally flawed as it implies that all papers have equal worth. It would, for example, be insulting to an ecologist whose research required years of painstaking fieldwork if during an assessment exercise their paper were given the same weight as a paper which applies a standard technique to a problem of little interest and which needed only a few days' effort overall. Different subdisciplines can have vastly dissimilar typical publication rates. In some fields the top researchers produce around one paper a year – for example, while Andrew Wiles worked for 8 years towards a result that led to

the proof of Fermat's Last Theorem he published just 2 papers. At the opposite extreme, the number theorist Paul Erdős wrote or co-authored 1475 papers (all of them substantial) during his life[1], and from 2005–8 the now retired Editor-in-Chief of the Elsevier journal *Chaos, Solitons, and Fractals*[2] managed around one per week on his theories of high-energy physics.

There is also no general relation between the quality of a researcher and the publication count. One might be tempted to suspect an inverse correlation above a certain threshold – a large number of papers might suggest that the work being published is relatively trivial – but there are also a number of internationally outstanding scientists who publish a lot of high-quality papers.

In assessing the research activity of departments or institutions, if publication counts are to be used, one should of course consider the number of publications per faculty member rather than the publication totals. However, in places where not all the faculty conduct research, an essential additional indicator would simply be the proportion of staff members active in research. A department where most of the staff publish at a moderate rate is surely more healthy than one where only a minority are active but with a similar number of papers per head as a result of the production of a multitude of mediocre articles.

Journal impact factor

A reasonable, but not faultless, method of gauging the value of a paper to the scientific community within a particular field is to look at the number of citations and compare it to that of other papers in the field. This can only be done a number of years later, after the paper has accumulated most of the citations it is going to receive. To assess the worth of recent papers, the most commonly used approach is to look at the journal impact factor (JIF). The JIF is defined as the number of citations during a particular year to journal articles within the previous n years divided by the total number of research articles in the journal during those n years, where usually $n = 2$. It is therefore a measure of the average citation rate (number of citations of an article per year). The numerous problems with evaluation using the JIF have long been recognized (for a detailed critique see Ref. 3), and various alternative metrics that lack some of these shortcomings have been proposed[4]. However, the fundamental problem with using the JIF or its alternatives is that one is of course assessing the journal and not the paper itself. The distribution of citation rates of articles in a given journal is highly skewed – some articles are cited a lot, which pulls up the JIF, whereas a significant proportion are never cited. As a result, only a small proportion of articles in a journal have a citation rate close to the JIF. One might think that using the JIF would be appropriate for collectively assessing a large group of publications such as from a university, but even on averaging over an entire country, the JIF has been found to differ significantly from the actual citation rate[3].

Important new results are normally submitted to prestigious journals but the prestige of a journal in a field does not always equate to its standing in terms of the JIF[5]. The JIF also does not always reflect other desirable aspects such as volume of readership and quality of writing; *American Mathematical Monthly*, a periodical one would expect to find in the coffee rooms of all respectable mathematics departments, has a relatively low JIF but exacting editing. Furthermore, some respected authors stick to their favourite journals for decades even though the JIF and/or prestige may wane. They may also eschew publishing in the top journals due to publication charges and other reasons[3]. Publishing in a highly cited journal does not lead to an article being cited more often than it would be in a low JIF journal[3].

Citation counts and the h-index

Having raised objections to basing assessment on the journal rather than the article, we turn to the fairer but less immediate approach of citation counts. It would seem reasonable to judge a paper which has never been cited after many years to be at most of marginal interest. However, for papers that have been cited more than a few times, there is little correlation between the number of citations a paper receives and its long-term significance as judged by experts[6]. This is partly due to the fact that flawed studies tend to get cited a lot as do topics that are currently in vogue[7].

As a simple modification, self-citations are sometimes excluded in citation counts. This is a little unfair if the self-citations are wholly justified. Some studies have also looked at the positions of citations in research articles[8]. The most crucial citations tend to occur in the methods section, and also in the discussion, provided it is not simply a list of comparisons with other studies. Perhaps a more revealing citation count could be obtained by giving a lower weight to citations from the introduction.

A convenient and rather telling statistic which is gaining popularity is the h-index[9]. An individual has an h-index of h if h of their articles have received at least h citations and the remainder have at most h citations. It measures the cumulative achievement of a scientist, and its rate of change, provided the scientist

remains productive, is a measure of their ability. As the originator of this statistic was the first to admit, it has a number of drawbacks[9]. Modified versions of the h-index omitting contributions from self-citation and taking into account multi-authored papers have since been proposed[10], as have other indices which circumvent the problem of researchers with a small number of highly cited papers ending up with a low score[11,12]. As with the other statistics, it is unfair to use it to compare researchers from fields with differing typical citation or publication rates.

NUMBERS GAMES

There are various ways the less ethical can artificially increase their scores. First, to obtain a higher publication count, publish more papers! We will not dwell on the most obvious ways of doing this which are also the worst types of fraud a scientist can commit – blatant plagiarism, self-plagiarism (submitting the same manuscript to more than one journal), and falsifying data. Such acts will be uncovered eventually and those involved blacklisted. Milder forms of plagiarism, however, are increasing and seem to be going unchecked[5]. As a result of inattentive referees and genuine or feigned ignorance of the literature by the authors, established results are being reclaimed as novel. Some authors simultaneously submit superficially differing manuscripts to rival journals but the main results they contain are essentially the same.

Additional papers can be generated without extra research by presenting the results of a study in a number of smaller papers. The papers need not necessarily be particularly brief – long introductions and discussions along with regurgitation of methods or reasoning given elsewhere can disguise the paucity of original results.

For the high-status researcher lacking in scruples, an extra publication is easily gained by requesting co-authorship on a subordinate's paper. This would be in return for a favour inversely related to the difference in status.

After an initial study leads to a publication, it is generally straightforward to carry out a series of similar studies which will in turn give rise to a series of articles of diminishing interest. In the experimental sciences, the same procedures are applied to various materials or organisms. For computational or theoretical studies, solving the next problem by mechanically applying exactly the same method as for the last can often be accomplished much more quickly. If the results are correct, make sense, and are obtained elegantly, they are at least of some use. But sadly, in a number of physics and applied mathematics journals in recent years there has been a plethora of papers attempting to solve problems using inefficient or inappropriate methods that lead to solutions that are either trivial or nonsensical (see Ref. 13 and references therein for a review). Obtaining solutions of nonlinear equations is another popular theme for easily produced publications. Solutions often turn out to be known solutions which have not been simplified or are just incorrect and the authors and referees have not taken the trouble to check[14]. Physically irrelevant solutions are also often presented without justification.

Boosting your citation count can be accomplished by citing your own work whenever possible, but it is better if you can obtain citations from others. You can cite other people's papers a lot in the hope that they will reciprocate. Referees sometimes encourage authors to cite the referee's own work, and naturally enough the authors dare not refuse. In some cases the suggestion is justified – the authors may have unwittingly missed a connection to a relevant paper – but in other cases, particularly when the referee suggests citing vaguely related references in the introduction, no improvement to the paper results.

Healthy numbers of citations to your work can be obtained if you are a member of a 'citation clique' – a group of authors who excessively cite each others' papers. These cliques tend to publish the types of easily-produced papers of suspect worth mentioned earlier. A further benefit of clique membership is that editors often pick referees from the list of citations, and clique members are likely to approve of each others' work. If you are not a member of a citation clique then why not create one of your own? Just find some unproductive authors who are willing to publish your work as their own. According to a blog about a now notorious author, a high-school maths teacher earned her PhD by allowing her name to be used.

Having more citations will naturally help one's h-index to a varying extent. The most efficient way to boost the index is to ensure that the $(h+1)$th most cited paper receives more citations[9].

With such pressure to publish in high JIF journals, questionable behaviour might also be occurring among more respectable authors. A referee may exaggerate the importance of a run-of-the-mill manuscript in the hope that the favour will be returned. The review process is of course blind, but in a narrow field the referee knows the authors are likely to guess their identity. This is possibly a contributing factor in the decision made last year by *Physical Review Letters* (generally regarded as the most prestigious physics journal) to try to raise standards by, among other things, requiring referees 'to support favourable rec-

ommendations with substantive reasons to publish.'

The JIF can be manipulated by both authors and editors. If each paper makes one extra citation to an article in the journal from the last 2 years then the JIF increases by 1. On returning the referees' comments to the author, some editors suggest adding a few recent citations to their journal 'to develop themes of research within the journal.' Authors who are concerned about the JIF of the journals in which they publish regularly may decide to do this of their own accord and hope other authors will do the same to give their favourite journal a helping hand. Ref. 15 contains a somewhat extreme example of needless citations; the third sentence of the introduction contains 19 citations to the journal in which it is published. The citations are, incidentally, all authored by the journal editor and have only a glancing relevance to the substance of the paper. Editors can also base their decision to publish on how much an article is likely to be cited rather than on its quality, and one editor, fearing the implications of a low JIF, sought out highly cited authors and encouraged them to submit to his journal[6].

THE DAMAGE TO SCIENTIFIC PROGRESS

The above games which some people are drawn into playing by rewards for high scores are of no to help to science, tarnish its image, and in many cases slow its progress. A preoccupation with publication statistics is causing some researchers (particularly young ones) to become too obsessed with the idea of publishing papers, the end of the scientific process, and the true goal of obtaining new interesting or useful knowledge is viewed as a bonus if it happens to occur. Articles start to become vehicles for engaging in numbers games, and as long as it can pass the referees, there is then no feeling that a manuscript should be particularly carefully written. Publishing should be regarded as a service to other scientists, engineers, administrators, or even politicians, informing them, in the most lucid and succinct manner possible, of results that are of interest or use.

From the evident lack of proper inspection by referees and editors, clearly many journals are straining under the weight of the number of publications. Obvious errors propagate from one paper to the next. Articles rife with grammatical errors are almost the norm in some journals. Having a large number of papers to sort through is also time-consuming for others in the field. It is of course more convenient for fellow scientists if one's study is disseminated in a single article rather than several. And with the multitude of mediocre articles in some journals, the few good papers are more likely to be overlooked.

More severe damage results directly from the assessing of studies using crude statistics. If numbers of publications rather than the utility of the results are regarded as a measure of success, some researchers will tend to choose topics which are guaranteed to produce results quickly, rather than more interesting and challenging problems. Also, since funding is limited, the more worthy studies will end up having a smaller share of the resources.

A related issue concerns doctoral degrees. These should be awarded for research that significantly adds to the existing body of knowledge. However, the minimum requirement of one international publication by some universities is slipping into being regarded as the sufficient requirement with the result that candidates can pass on the basis of rather insubstantial results.

Assessing using the JIF can also result in the marginalization of important work. In medicine, for example, much damage has already been done in the UK where grant-awarding bodies have judged work by the JIF and as a result institutes have prioritized studies that are likely to result in publications in journals with a high JIF to the detriment of other equally vital studies that tend to appear in lower JIF journals[6].

Finally, basing promotions and awards on crude statistics may favour those who are adept at playing numbers games. Such people are likely to have questionable integrity, and as a result may cause problems for the reputable researchers beneath them.

RECOMMENDATIONS

A solution to the problems outlined would require reducing the incentives for people to play numbers games and thwarting attempts at manipulation. So what are the incentives and their rationale? We have touched on some already – the award of grants or positions based on scores rather than quality. It is done this way because it is easy. Some universities give financial incentives to their departments or individuals within them to maximize their scores. The main reason seems to be to help the ranking in university league tables which is partly based on crude publication statistics. But who are these rankings for? They are an average and so say nothing about each department. Prospective undergraduates would want to know how good the teaching and possibly research is in their major. Prospective graduate students, researchers, and research funding agencies would only be interested in the activity of a particular research group. Administrators should ignore the league tables and instead devote their energies to creating an environment in which quality science flourishes.

The publishers themselves also heavily promote the use of these statistics. Aside from prominently displaying the JIF on journal homepages, they give awards to authors based on statistics and even market this as a service. The problem is that these numbers do not measure quality. Sometimes the awards are appropriate – the work has been cited or downloaded a lot because it really is a useful contribution. But in other cases, the seminal work is a prototype for producing endless papers of dubious value.

In the face of fickle authors chasing high JIF journals, the standard advice for editors is still to always try to publish the best articles possible[6]. Care must be taken when choosing referees to avoid sub-standard papers being accepted due to recommendations from members of a citation clique. Authors who appear to be submitting a string of manuscripts on similar subjects should be encouraged to combine them. They can also be asked to declare the existence of similar papers they have published. Not carrying out some of these measures could well cause the JIF to increase – the later partial result will cite the one before, and once a citation clique is established, their articles will soon be receiving multiple citations. However, resisting the temptation to ignore games that may increase the JIF will reap rewards as a result of respectable authors choosing to submit articles after observing that the journal is maintaining standards.

If one does have to rely on statistics, the more sophisticated the measure, the more difficult it will be to manipulate. However, one suspects that whatever new statistical method of evaluation is used, people will learn the tricks to maximize their score. When using statistics for a preliminary assessment of a collection of individuals in particular, one should also look at the variation of the scores over time. Rapid increases, beyond what could be expected from statistical fluctuations or a steady improvement, are quite likely to be due to gaming.

How does one carry out assessment without statistics? Do the Nobel committee look at the number of papers or citations? No – they look at the significance of the science. Similarly, a candidate for promotion or award could be asked to describe the significance of what they have done. The report could be treated in the same way as a submitted manuscript or a DSc application and be assessed by referees in the field (which may well be from abroad). As with referee reports, the committee can easily judge whether or not the assessment is thorough and unbiased. Another key indicator is whether the candidate has been an expenses paid invited speaker at a (bona fide) international conference. Such scientists are evidently highly regarded by their peers.

In the first instance, statistics can be a useful guide and possibly the only option if resources are not available for assessing people by examining reports from referees. However, when the assignment of a large grant or a significant appointment is being considered, it is surely imperative to obtain a sound assessment rather than rely on statistics that are inherently unfair and easy to manipulate. Not doing so means that high quality science is not favoured over low quality science and the former suffers as a result.

Acknowledgements: I am indebted to friends and colleagues for discussions that played a major role in the inspiration behind and choice of content of this article, and for their critical reading of the manuscript.

REFERENCES

1. Hoffman P (1998) *The Man Who Loved Only Numbers*, Fourth Estate, London.
2. Schiermeier Q (2008) Self-publishing editor set to retire. *Nature* **456**, 432.
3. Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ* **314**, 497–502.
4. Bollen J, Rodriquez MA, Van de Sompel H (2006) Journal status. *Scientometrics* **69**, 669–87.
5. Arnold DN (2009) Integrity under attack: The state of scholarly publishing. *SIAM News* **42**(10).
6. Brown H (2007) How impact factors changed medical publishing—and science. *BMJ* **334**, 561–4.
7. Spier RE (2009) On the ethics of using citation indices in evaluations. *Sci Eng Ethics* **15**, 1–2.
8. Sombatsompop N, Kositchaiyong A, Markpin T, Inrit S (2006) Scientific evaluations of citation quality of international research articles in the SCI database: Thailand case study. *Scientometrics* **66**, 521–35.
9. Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* **102**, 16569–72.
10. Schreiber M (2009) The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. *Ann Physik* **18**, 607–21.
11. Egghe L (2006) Theory and practise of the g-index. *Scientometrics* **69**, 131–52.
12. Vinkler P (2009) The π-index: a new indicator for assessing scientific impact. *J Inform Sci* **35**, 602–12.
13. Fernández FM (2009) On some approximate methods for nonlinear models. *Appl Math Comput* **215**, 168–74.
14. Kudryashov NA (2009) Seven common errors in finding exact solutions of nonlinear differential equations. *Comm Nonlinear Sci Numer Simulat* **14**, 3507–29.
15. Falcón S, Plaza Á (2007) The k-Fibonacci sequence and the Pascal 2-triangle. *Chaos Soliton Fract* **33**, 38–49.