ON THE MEASUREMENT RESEARCH PERFORMANCE IN MATHEMATICS AND STATISTICS[1]

Peter Hall
The University of Melbourne

We live in an age where the notion that almost anything can, and should, be quantified and analysed, at an elementary and accessible level, is rapidly gaining adherents. In some countries the quantification of past research performance, and the prediction of performance in the future, have become major goals of university research managers. The principal objective of quantification is its use as a management tool, creating an imperative that the means of quantification be closely scrutinised.

At least five numerical measures are, or have been, used frequently to quantify research performance: (i)ñumber of research papers published, (ii)ñumber of pages published in research papers, (iii)ñumber of citations received, (iv) "impact factors" of journals where publication takes place, and (v) usage data on published papers. Item (ii) is sometimes normalised, e.g. for words per page, and (iii) and (iv) can also involve crude standardisation, for example to correct for relative citation rates in different fields. Data of type (v) tend to be available only from primary sources (e.g. publishers and journal archivers), not secondary sources that address a broad range of journals. This restricts the opportunities for easy comparison, particularly in multidisciplinary fields.

Citation data for an individual can be treated in a variety of different ways. These include: the total number of citations, the average number of citations per paper, the number of papers with at least x citations, and the largest value of x for which there are at least x papers with x citations; see, for example, Hirsch (2005). New methodologies are constantly under development (e.g. Sidiropoulos, 2006).

Ease of accessibility of data is a major motivator of some approaches. Thus, although the use of publication-rate data, such as those in categories (i) and (ii), can be criticised fairly on the grounds that it addresses quantity rather than quality, that approach was employed widely until relatively recently, when citation data became easily accessible via the internet. While citation and usage data are also widely criticised, and arguments against them are made frequently (e.g. Ewing, 2004, 2006; Monastersky, 2005), their ready availability today makes them attractive to research managers.

Publication-rate data for very highly-performing researchers in the mathematical sciences show particularly wide variation. In some areas of theoretical mathematics, for example number theory, it is not uncommon for career-long publication rates to be less one paper per year, with runs of several years without publication while especially difficult problems are tackled. This applies even to acknowledged international high-achievers in the field.

However, in other areas of the mathematical sciences publication rates can be substantially higher. This variability reflects a variety of factors, including different ways in which researchers work, disparate amounts of time needed in different fields to obtain significant new results, and cultural differences between areas as to what constitutes a "significant advance."

In the face of difficulties using information on publication rates to assess research performance, university research managers are turning increasingly to citation data. Here several intrinsic, but subtle, statistical issues have a substantial bearing on interpretation. In particular, the distribution of citation data is very heavy-tailed; that is, a relatively large proportion of the distribution is concentrated among quite high values. Therefore it is unsurprising to learn that the mean of the distribution (for instance, of the distribution of the average number of citations of a given paper in a particular journal during a given time period), is almost always larger than, and can be substantially greater than, the median. Similar remarks apply to the totals that are used to compute means, for example to the total number of citations received by a paper in a given period. This has important implications

[1] This note is based in part on the article, "Measuring research performance in the mathematical sciences in Australian universities," published in the Australian Mathematical Society Gazette vol. 34, No. 1, 2007.

for the use and interpretation of citation data.

For example, since most citation indices (e.g. impact factors) are means rather than medians, their values can be altered dramatically by including, or omitting, a single research paper in the calculations. This is one reason why impact factors tend to fluctuate significantly from one year to another. Another reason is that impact factors often rely on relatively narrow time windows, and so, at least in the mathematical sciences, tend to be based on relatively small amounts of data.

The width of the citation window is a contentious issue when gathering and interpreting citation data. Theoretical mathematicians, for many of whom the solving of important, years-old problems is a mark of singular achievement, naturally regard relatively wide windows (at least 10 years) as a major desirable feature of approaches to citation analysis. However, if the window is as wide as a decade then the period over which the mathematician's performance is supposedly being assessed is arguably wider still, and that is not necessary desired by those doing the assessment. (The serious biases inherent in using narrow citation windows to assess papers with relatively long citation lives have been studied; see e.g. Vanclay, 2006.) Moreover, researchers in other disciplines, with fast-moving research frontiers, often favour relatively narrow windows. The latter tend to prevail.

As a result, mathematicians and statisticians generally find that they are judged using an unreasonably narrow citation window, which almost inevitably obscures the real degree of interest in, and impact of, their work. The two-to-three-year impact factors for major mathematics and statistics journals are typically about 1 or 2. That is, on average a paper is cited once or twice in the year of publication or in the subsequent two years. This compares poorly with the two-to-three-year impact factors of approximately 30 for journals such as *Nature* and *Science*, but of course does not indicate any intrinsic inferiority of research in the mathematical sciences. Rather, it is the result of different citation cultures in different fields of science. For reasons such as these there is an extensive literature arguing against using impact factors to evaluate research (e.g. Moed et al., 1996; Jennings, 1998; Seglen, 1997; Jasco, 2001; Whitehouse, 2001; Simkin and Roychowdhury, 2003).

The speed of publication of mathematics and statistics papers also has significant bearing on the use of citation windows. Publication in major international journals, which often have long lead times, is itself a validation of mathematical and statistical work. For some mathematical scientists, such publication is almost as much a goal as the research itself. However, papers in major mathematics and statistics journals often take longer from submission to publication than an average citation window takes to run its course. This inevitably challenges conventional interpretations of citation data.

Other issues related to the reliability of citation data include the fact that those data do not identify the reasons for citation, or disclose which authors of a multi-authored paper are responsible for different aspects of the work. In statistics, reasons for high citation rates can include the fact that a useful dataset was included in the paper, or that helpful settings for simulation studies were suggested. These contributions may not bear on the actual merit of the research.

Moreover, citation cultures can vary widely even within a single discipline, such as the mathematical sciences. This leads to very different "natural" citation rates for people working in different parts of the same field. The Stanford statistical scientist David Donoho, interviewed when he was the "most highly-cited mathematician for work in the period 1994–2004" (Leong, 2004), put it thus:

> Statisticians do very well compared to mathematicians in citation counts. Among the top ten most-cited mathematical scientists currently, all of them are statisticians. There's a clear reason: statisticians do things used by many people; in contrast, few people outside of mathematics can directly cite cutting-edge work in mathematics. Consider [Andrew] Wiles' proof of Fermat's Last Theorem. It's a brilliant achievement of the human mind but not directly useful outside of math. It gets a lot of popular attention, but not very many citations in the scientific literature. Statisticians explicitly design tools that are useful for scientists and engineers, everywhere, every day. So citation counts for statisticians follow from the nature of our discipline.

Donoho also commented on ways in which one can enhance one's visibility in citation counts:

> A very specific publishing discipline can enhance citation counts: Reproducible Research. You use the internet to publish the data and computer programs that generate your results. I learned this discipline from the seismologist Jon Claerbout. This increases your citation counts, for a very simple reason. When researchers developing new methods look for ways to show off their new methods they'll naturally want to make comparisons with previous approaches. By publishing your data and methods, you make it easy for later researchers to compare with you, and then they cite you.

Remarks such as these inevitably provoke the question of the relationships among impact, influence and quality in research. Research can have substantial impact (for example, through enabling other researchers to "show off their new methods," as Donoho put it), and give rise to large numbers of citations, without significantly altering the intrinsic directions taken by future research, and therefore without having much influence in that sense. In much the same way, a movie can enjoy substantial box-office success without having a major influence on movie-making.

Occasionally, the order of authors on a paper is proposed by research managers as a measure of the relative importance of individual contributions. However, in much of mathematics the order of authors is almost religiously alphabetical. (This is not so true of statistics.) The suggestion by some research managers that mathematical scientists change these practices, by ordering author names so as to reflect respective contributions to papers, or by altering their culture of publication and citation (e.g. by publishing and citing more frequently), or by submitting only to journals where lead-times to publication are measured in weeks or months rather than years, fail to take account of the fact that only a small fraction of the international mathematics literature originates in countries such as Spain and Australia. Profound cultural change cannot be brought about by doing things differently in just one country.

In summary, research performance metrics, such as those based on publication rates or citation or usage data, often do not measure the research attributes that it is claimed they do. They lack comparability, even from area to area within a single discipline, such as the mathematical sciences, let alone from one discipline to another. So-called correction factors fail to compensate adequately for the marked inhomogeneity of citation cultures, for example those in applied and theoretical statistics. In the absence of reliable and accepted ways of correcting for the problems discussed above, the use of research performance metrics is inevitably a crude and unreliable way of assessing actual research performance. Careful, scholarly assessment of research by peers is usually to be preferred.

Acknowledgements. Helpful discussion with Gus Lehrer and Alf van der Poorten is gratefully acknowledged.

Referencias

[1] Ewing, J. (2004). Why the AMS does not provide journal usage statistics. *Amer. Math. Soc. documents*, `http://www.ams.org/ewing/Documents/NoStatistics-43.pdf`

[2] Ewing, J. (2006). Measuring journals. *Not. Amer. Math. Soc.*, **53**(9). Available at: `http://www.ams.org/notices/200609/comm-ewing.pdf#search=%22ewing%20%22measuring%20journals%22%22`

[3] Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Manuscript.

[4] Jacso, P. (2001). A deficiency in the algorithm for calculating the impact factor of scholarly journals: The journal impact factor. *Cortex*, **37**, 590–594.

[5] Jennings, C. (1998). Citation data: the wrong impact? *Nature Neuroscience*, **1**, 641–642.

[6] Monastersky, R. (2005). The number that's devouring science. *Chronicle Higher Ed.* `http://chronicle.com/free/v52/i08/08a01201.htm`

[7] Leong, Y.K. (2004). David Donoho: Sparse data, beautiful mine. Excerpts of interview of David Donoho by Y.K. Leong. *News-*

letter of Institute for Mathematical Sciences, Issue No. 5 for 2004. National University of Singapore. http://www.ims.nus.edu.sg/imprints/interviews/ DavidDonoho.pdf

[8] Moed H.F. and Van Leeuwen, T.N. (1996). Impact factors can mislead. *Nature*, **381**, 186.

[9] Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *Brit. Med. J.*, **314**, 498–502. http://www.bmj.com/cgi/content/full/314/7079/497

[10] Sidiropoulos A., Katsaros D. and Manolopoulos Y. (2006). Generalized h-index for disclosing latent facts in citation networks. Manuscript.

[11] Simkin, M.V. and Roychowdhury, V.P. (2003). Read before you cite! *Complex Syst.*, **14**, 269–274.

[12] Vanclay, J.K. (2006). Bias in the journal impact factor. *Scientometrics*, to appear. http://arxiv.org/pdf/cs.DL/0612091.

[13] Whitehouse, G.C. (2001). Citation rates and impact factors: should they matter? *Brit. J Radiology*, **74**, 1-3.