Measuring research performance in the mathematical sciences in Australian universities

Peter Hall

The paper below was prepared in response to a request, in September 2006, from an Australian Mathematical Society committee. It was intended to provide background to general discussion of research-performance metrics in the mathematical sciences. At the time the paper was solicited it was still uncertain whether or when the Australian Government would implement the Research Quality Framework (RQF) to assess publicly funded research. But despite this uncertainty mathematical scientists in universities were coming under pressure to agree to metrics that could be used in submissions to the RQF, if it were to go ahead.

The Productivity Commission, in its Draft Research Report released on 2 November 2006, argued that 'it is too early to make a final decision about implementation of the RQF', and stated that the RQF's 'adoption should be delayed'. However, 12 days later the Australian Government endorsed the RQF, and announced that 'preparatory work and trialling will continue in 2007, with data collection in 2008 and funding implementation in 2009'.

Since that time, members of the Society have begun developing approaches and advice that might be used to assist mathematical sciences departments to prepare submissions to the RQF. An informal committee has been set up for this purpose, and can be contacted through Professor Peter Taylor (pgt@ms.unimelb.edu.au) at The University of Melbourne. Publication of the paper below stems from a request that it be made public so that it might be used in connection with that work.

Measuring research performance in the mathematical sciences in Australian universities

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 Australian universities the quantification of past research performance, and the prediction of performance in the future, have become major goals of 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) number of research papers published, (ii) number of pages published in research papers, (iii) number of citations received, (iv) 'impact factors' of journals where publication takes place, and (v) usage data on published papers. Item (ii) is sometimes normalised, for example 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 (such as 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, [3]. New methodologies are constantly under development (for example [6]).

The degree of accessibility of data is a major motivator of different 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 (for example [1], [2], [4]), their ready availability today makes them attractive.

Publication-rate data for 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 than 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, such as Fields Medallists.

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'.

Graduate student numbers likewise show a remarkable degree of variability from one area to another in mathematics, reflecting both the amount of scholarly preparation needed and the level of demand for graduates. Several of Australia's most highly respected theoretical mathematicians have had relatively few graduate students during their careers. Reasons include the fact that, in some areas, the levels of knowledge required before embarking on PhD-level research are so great as to discourage all but the most able and dedicated students. On the other hand, in other areas of mathematics, including some where research frontiers change rapidly, significant results can be achieved relatively quickly, using tools that sometimes can be acquired in advanced undergraduate courses. Here students do not need to devote long periods of time to preparation, and tend to be more inclined to undertake graduate work.

In still other parts of the mathematical sciences, salaries and working conditions outside the university sector are so enticing that it can be very difficult to attract good graduate students. Ironically, these areas tend to be of substantial, immediate strategic importance to the nation, and so a deficit of graduate students can occur in precisely those areas where relatively large numbers would be desirable. For all these reasons, the numbers of PhD students supervised by individual mathematical scientists are very poor indicators of relative levels of research activity.

In the face of difficulties using information on publication rates and graduate student numbers to assess research performance, research managers in Australian universities 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 or 'average value' 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 (or 'middle value'). 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 for the use and interpretation of citation data.

For example, since most citation indices (for example 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 are often based 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. 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 necessarily desired by those doing the assessment. Moreover, researchers in other disciplines, with fast-moving research frontiers, often favour relatively narrow windows. The latter tend to prevail.

As a result, mathematicians 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 some of the most prestigious mathematics journals, especially those in theoretical mathematics, are typically about 1. That is, on average a mathematics paper is cited approximately once 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.

The speed of publication of mathematical results 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 work. For many mathematicians, such publication is almost as much a goal as the research itself. However, papers in pre-eminent mathematics 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 applied areas, 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' [5], 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 the mathematical sciences the order of authors is almost religiously alphabetical. The suggestion by some managers that Australian mathematicians change these practices, by ordering author names so as to reflect respective contributions to papers, or by altering their culture of publication and citation (for example, 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 Australia. Profound cultural change cannot be brought about by doing things differently in Australia; the actions mentioned above would lead only to a substantial diminution of the international reputation of Australian mathematics.

The Australian Government Department of Education, Science and Training commonly refers to the metrics which inform its Research Training Scheme, Institutional Grants Scheme, and Research Infrastructure Block Grants Scheme, as indicators of 'research performance'.

These metrics, based on numbers of papers published, numbers of graduate students trained, and numbers of dollars of research funding brought into a university (through either competitive grants or commercial work), are applied to all universities. They lead to research budgets for individual institutions. Research managers, eager to increase the money flowing into their respective universities, sometimes carry these metrics right down to the level of individual mathematical scientists, pressuring them to increase what the managers term the mathematicians' 'research performance' (that is, to increase the university overhead income that results from the mathematicians' research), or suffer the consequences. This is a tawdry approach to research management.

In summary, research performance metrics, such as those based on publication rates, numbers of graduate students, 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.

Acknowledgements

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

References

[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**, No. 9. Viewed 20 January 2007. http://www.ams.orgnotices/200609/comm-ewing.pdf

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

[4] Monastersky, R. (2005). The number that's devouring science. *Chronicle Higher Ed.* Viewed 20 January 2007. http://chronicle.com/free/v52/i08/08a01201.htm

[5] Leong, Y.K. (2004). David Donoho: Sparse data, beautiful mine. Excerpts of interview of David Donoho by Y.K. Leong. *Newsletter of Institute for Mathematical Sciences*, Issue No. 5 for 2004. National University of Singapore. Viewed 20 January 2007. http://www.ims.nus.edu.sg/imprints/interviews/DavidDonoho.pdf

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

School of Mathematics & Statistics, The University of Melbourne, VIC 3010
*Email:*halpstat@ms.unimelb.edu.au