Statistical Science
2009, Vol. 24, No. 1, 25–26
DOI: 10.1214/09-STS285D
Main article DOI: 10.1214/09-STS285

Comment: Citation Statistics

Peter Gavin Hall

Key words and phrases: Bibliometric analysis, bibliometric data, citation analysis, impact factor, journal ranking, research assessment.

I remember a US colleague commenting, in the mid 1980s, on the predilection of deans and other university managers for assessing academic statisticians' performance in terms of the numbers of papers they published. The managers, he said, "don't have many skills, but they can count." It's not clear whether the management science of assessing research performance in universities has advanced greatly in the intervening quarter century, but there are certainly more things to count than ever before, and there are increasingly sophisticated ways of doing the counting.

The paper by Adler, Ewing and Taylor is rightly critical of many of the practices, and arguments, that are based on counting citations. The authors are to be congratulated for producing a forthright and informative document, which is already being read by scientists in fields outside the mathematical sciences. For example, I mentioned the paper at a meeting of the executive of an Australian science body, and found that its very existence generated considerable interest. Even in fields where impact factors, h-factors and their brethren are more widely accepted than in mathematics or statistics, there is apprehension that the use of those numbers is getting out of hand, and that their implications are poorly understood.

The latter point should be of particular concern. We know, sometimes from bitter experience, of some of the statistical challenges of comparing journals or scientists on the basis of citation data—for example, the data can be very heavy-tailed, and there are vast differences in citation culture among different areas of science and technology. There are major differences even within probability and statistics. However, we have only rudimentary tools for quantifying this variation, and that means that we can provide only limited advice to people who are using citation data to assess the work of others, or who are themselves being assessed using those data.

Peter Gavin Hall is Professor of Statistics, University of Melbourne, Victoria, Melbourne, VIC 3010, Australia (e-mail: halpstat@ms.unimelb. edu.au).

Therefore, one of the conclusions we should draw from the study by Adler, Ewing and Taylor is that we need to know more. Perhaps, as statisticians, we could undertake a study, possibly funded in part by a grant awarding agency or our professional societies, into the nature of citation data, the information they contain, and the methods for analysing them if one must. This would possibly require the assistance of companies or organizations that gather such data, for example, Thomson Reuters and the American Mathematical Society. However, without a proper study of the data to determine its features and to develop guidelines for people who are inevitably going to use it, we are all in the dark. This includes the people who sell the data, those who use it to assess research performance and those of us whose performance is judged.

It should be mentioned, however, that too sharp a focus on citation analysis and performance rankings can lead almost inevitably to short- rather than long-term fostering of research excellence. For example, the appropriate time window for analyzing citation data in mathematics and statistics is often far longer than the two to three years found in most impact factor calculations; it can be more like 10–20 years. However, university managers typically object to that sort of window, not least because they wish to assess our performance over the last few years, not over the last decade or so. More generally, focusing sharply on citations to measure performance is not unlike ranking a movie in terms of its box-office receipts. There are many movies, and many research papers, that have a marked long-term impact through a complex process that is poorly represented by a simple average of naive criteria. Moreover, by relying on a formulaic approach to measuring performance we act to discourage the creative young men and women whom we want to take up research careers in statistical science. If they enjoyed being narrowly sized and measured by bean-counters, they'd most likely have chosen a different profession.

To illustrate some of the issues connected with citation analysis I should mention recent experiences in Australia with the use of citation data to assess research