

Editorial

Bibliometrics and a culture of measurement

Who knows whether it would have happened anyway, without Margaret Thatcher and a decade or two of public sector managerialism but, increasingly, aspects of our working and, supposedly, private lives are becoming subject to inspection and measurement. Just as new optical instruments and mathematical techniques ushered in the Enlightenment in Europe a few hundred years ago (Russell 1991), new technology today seems to have intensified the possibility for scrutiny and measurement as well as for parodies of measurement. We say parodies of measurement for three reasons. First, numbers, the 'canon of the Enlightenment' as Adorno and Horkheimer (1979) argued, attempt to make 'the dissimilar comparable by reducing it to abstract quantities'. Problems of assigning entities to categories and then counting them are not simply technical, of course, but involve ultimately subjective and politically interested decisions. Whose and which activities should be recorded? How should they be represented and how can they be translated into performance, reward and discipline? What forces shall be allowed to remain invisible? Parody, secondly, because various prominent measures, whether surgeon or school league tables, can fail to take enormous contextual factors into account, yet they are still heavily invested in by governments and, at least in the case of schools, consumed in daily newspapers. Parody, thirdly, because there is a tendency for those whose activities and characteristics are being measured to turn considerable attention towards the measure, rather than that quality supposedly being measured, in order to achieve some kind of reward, or, more usually in the public sector, stave off some penalty.

Research, like other areas of public life, is coming under increasing pressure

to justify the return on investment. In science as a whole there is increasing interest in comparative indicators of international performance. Within the highly visible policy field of health care in the United Kingdom (UK), the present government seems intent on strengthening audit culture and its grasp on measurement and performance for example by means of National Service Frameworks.

It is in this climate that bibliometrics has become attractive to research funders and policy makers, with its promise of some objective measure of output. Like other normalizing technologies, bibliometrics makes it possible to arrange individuals, departments, disciplines and journals into a graded array of productivity or quality.

Bibliometrics itself is not a new discipline and its development certainly precedes global recessions and oil crises that partly led to public sector stringency in the developed world. The Institute for Scientific Information (ISI) was founded in 1958 and, as a result of United States National Institutes of Health funding obtained in 1961, developed a database that eventually became the Science Citation Index. The Social Sciences Citation Index emerged 11 years later. It was not, however, until the early to mid-1990s that, in the UK, the Wellcome Trust developed its own Research Outputs Database (ROD). Based on ISI data, this is a database of citation of biomedical research papers originating in the UK whose funding acknowledgements and precise addresses have been added. Like so many social and political phenomena, part of the background to its conception was the apparent external pressure on funding bodies for increased accountability and internal pressure to target resources more effectively (PRISM 1995). Bibliometrics proceeds from the assumption of the equivalence of citation with usefulness to the scientific community. However, work on the payback from

research reveals what a complex area this is, particularly when trying to assess public impact (Buxton & Hanney 1998). In a recent study that we carried out in collaboration with the Wellcome Trust, conventional citation scores were not used at all.

Publishers, journal editors, and academics alike will have a keen interest in one particular bibliometric indicator, the impact factor. This is one of three measures devised by the ISI as a way of describing the pattern of citation to articles, which originally appeared in any particular journal. The impact factor is a way of rating and, as always, ranking journals in terms of their perceived importance among the community that they serve. If the level of citations to articles is considered as a curve, this generally rises sharply to a peak between 2 and 6 years after a paper's publication and then declines slowly over a much longer period. The impact factor is a measure of the relative size of the citation curve between 2 and 3 years after publication. It is calculated by dividing the number of current citations a journal receives to articles it published in the previous 2 years by the total number of articles published during those 2 years. However, despite being the most widely used measure, it is anomalous because the numerator refers to all publications (including editorials, news items and letters to the editor) in a journal, while the denominator includes only articles, notes, letters and reviews (Lewison 2001). In addition, if the window of measurement is widened from 2 to 5 years, there can be dramatic changes in the ranking of journals being examined. For example, in a study reported by Amin and Mabe (2000), 24 out of 30 chemistry journals changed rank, by as much as 11 places when the window was widened in this way. Typical citation factors vary greatly between different disciplines for reasons to do with differences in disciplinary practice. The mean impact factor for fundamental life

science journals is a little over 3, while for social sciences it is approximately 0.5 (Amin & Mabe 2000). Thus it is meaningless to compare journals from different disciplines by simply comparing impact factors.

Each year ISI publishes a journal citation report which it describes as a 'systematic, objective way to determine the relative importance of journals within their subject categories' and as ways for publishers to monitor competitors (Institute for Scientific Information 2001). Thus their reports have a high market value themselves. The report for 1999, for example, includes 42 journals under its nursing heading. The US journal *Nursing Research* had the highest impact factor of 1.090 and the *Journal of Advanced Nursing* was the highest ranked UK-based journal, 13th with an impact factor of 0.638. By contrast, the category psychology includes 107 journals with the highest having an impact factor of 7.790. Journal publishers may well be tempted to try to enhance their citation levels in any way they can, for example by encouraging authors to include citations to their own journals but it is difficult to stand on any moral high ground on this issue as most academics, at some point, will have developed the habit of including a few citations to their own work in papers they are writing, for similar reasons.

Reputation is self-fulfilling. A highly esteemed journal will attract the attentions of authors who wish their work to be assessed in the best light possible. As we mentioned above, in our recent bibliometric study of the outputs of nursing research from 1988 to 1995 (Rafferty *et al.* 2000; Traynor *et al.* 2001), our collaborator from the Wellcome Trust rejected conventional impact factors as a useful estimation of esteem and instead we contacted two panels, one of established nurse researchers and one of practice-based research leads and asked them to rate the importance and influence of a range of nursing journals. As our collaborator, Grant Lewison, later argued:

...in biomedicine, the object of research is not merely to accumulate citations, however, gratifying this may be for the individual researchers. It is rather to develop an understanding of the subject so that patients may

be given better treatment, or prevented from becoming ill or injured in the first place. Account should therefore be taken of how likely the papers in a given journal are to influence clinical practice when they are being evaluated. (Lewison 2001, p. 2)

In our study there was virtually no correlation between the estimations of users and researchers on the one hand and conventional impact measures on the other. Both our groups unanimously agreed that the *Journal of Advanced Nursing* was in the top category, however, informally, in our experience, groups of practising nurses have often named *JAN* as an example of an inaccessible and irrelevant forum for their own inquiries into practice problems. This strong difference of opinion may well be characteristic of a practice discipline with diverse membership and our guess would be that this divergence is not a characteristic of some other basic sciences where their communities would be more discrete and homogenous. We believe that the issue of research relevance for practice is far more complex than it is often made out to be. There is certainly a place for a research community to 'talk (more or less) to itself' to debate, develop, refine or reject theories, methods and tools that may form the basis of tomorrow's applied research. Assessing influence and usefulness is, in our view, even more difficult than Lewison is arguing. It may be all but impossible to understand how individual journals influence practitioner/manager/policy maker's actions and consciousness. In our study, when our panels were asked to give their estimation of how influential or important a range of journals were, they may well have reported on their reputations rather than on some more objective account of their actual influence.

Like many attempts to measure social phenomena, bibliometrics can give broad indications of differences and trends but more detailed measurements and comparisons may well turn out to be spurious for some of the reasons we have mentioned. It would be wrong to understand bibliometrics simply as a technology of control. It has more creative possibilities and can help us map the intellectual growth of a discipline and help participants to understand patterns of collabor-

ation and funding. A movement from number 11 to number 10 in journal citation rankings may provide a boost in morale for the publisher concerned but whether it represents any useful change in quality or impact is open to question.

Michael Traynor
Centre for Policy in Nursing Research,
London School of Hygiene and
Tropical Medicine,
UK

Anne Marie Rafferty
Centre for Policy in Nursing Research,
London School of Hygiene and
Tropical Medicine,
UK

References

- Adorno T. & Horkheimer A. (1979) *Dialectic of Enlightenment*. Verso, London.
- Amin M. & Mabe M. (2000) *Impact Factors. Use and Abuse*. Elsevier Science, Oxford.
- Buxton M. & Hanney S. (1998) 'Evaluating the NHS Research & Development programme: will the programme give value for money?' *Journal of the Royal Society of Medicine* 91(S35), 2-6.
- Institute for Scientific Information (2001) *Journal Citation Reports on the web*: <http://sunweb.isinet.com/isi/products/citation/jcr/jcrweb/index.html>. Last accessed 19.6.2001.
- Lewison G. (2001) *Researchers' and Users' Perceptions of the Relative Standing of Biomedical Papers in Different Journals*. Scientometrics and Informetrics, Sydney, Australia.
- PRISM (1995) *Briefing Note on Evaluation Service: Investigating the Effectiveness of Different Schemes for Providing Funding*. Unit for Policy Research in Science and Medicine, The Wellcome Trust, London.
- Rafferty A.-M., Traynor M. & Lewison G. (2000) *Measuring the Outputs of Nursing R & D: a Third Working Paper*. Centre for Policy in Nursing Research, London School of Hygiene and Tropical Medicine, London.
- Russell B. (1991) *History of Western Philosophy*. Routledge, London.
- Traynor M., Rafferty A.M. & Lewison G. (2001) Endogenous and exogenous research? Findings from a bibliometric study of UK nursing research. *Journal of Advanced Nursing* 34, 212-222.