

Beyond *SCI* citations – New ways to evaluate research

Grant Lewison

City University, London EC1V 0HB, UK

Keywords: Eugene Garfield, linkages, *SCI* citation.

The creation of the *Science Citation Index (SCI)* back in the 1960s by Eugene Garfield^{1,2} was a major achievement for scholarship and the retrieval of relevant literature, but also for the practice of research evaluation. Over the years, a wealth of literature has been generated on the compilations of these ‘conventional’ citation scores for individual research papers, and aggregate values for journals, and the statistics based on them. There is often dispute on how much weight should be attached to one research paper citing another in its list of references^{3,4} and there has also been criticism of the inappropriate use of citation scores to compare outputs in different subjects⁵. Nevertheless, the use of citation data can add a welcome objective contribution to the subjective research evaluation performed by peer-review groups^{6,7}.

However there remains the limitation that citations by researchers only represent the first of many steps by which research is put to use by society, either to generate new and improved industrial products and processes, or to promote public goods such as better health or a cleaner environment. Figure 1 shows that, for biomedical research, there is a complex web and many links between the conduct of research and the ultimate goals of better patient care and less illness. In this diagram, there are many important nodes, such as government policy and public attitudes, which are often neglected in the research evaluation process. This paper seeks to trace some of these linkages and to provide means to quantify them and evaluate the utility of the underlying research through the medium of citations, but citations from documents other than research papers in the peer-reviewed journals processed for the *SCI*.

It should be emphasized that research evaluation is not simply a means whereby the outputs of an entity – a country, an institution or a research group – can be characterized by a single indicator. Research is a complex process with many routes to influence the external world, and so a range of indicators is needed to put a value upon it⁸. Not all of them will be giving the same message, and the choice of which ones to select will depend on the type of research being evaluated⁹. The indicators are in some ways comparable to the parameters that describe a car’s

performance. The ones appropriate to a sports car will differ from those for a family saloon; moreover different prospective purchasers will be influenced by different parameters – top speed, fuel consumption, luggage capacity, seating comfort, resale value. *SCI* citations and journal impact factors could be considered as analogous to acceleration and top speed, but there are many other parameters of car performance and there are many indicators of research utility.

There are several important features of these new tools for research evaluation:

- They are national – or even regional – rather than international, and so will require corresponding data-gathering exercises. These should be carried out using agreed norms and protocols so that data can be exchanged with colleagues in other countries.
- They will nevertheless provide information not only on national research outputs, but on those from other countries. (An example is given below.) The more countries take part in the exercises, the richer will be the resource for research evaluation.
- They apply very differently to different types of research: applied/clinical or basic, so the selection of indicators must be appropriate for the research being evaluated.
- Most research papers will not be cited at all in these other documents, so the new indicators must be applied to large numbers of papers (at least several hundred) if statistical comparisons are to have any significance.
- Some of these citation indicators may reach a peak much earlier than conventional citations (whose peak occurrence is usually 2–3 years after publication), or much later, and the time window allowed for data gathering should take account of this.

Whereas the norms for conventional citations (the ones in the *SCI*) are reasonably well developed, even if not always appropriately used, the norms for these new indicators are much less known, except for one of them, namely citations on patents. These have been studied in some detail by CHI Research Inc. in the USA^{10,11}. It appears that nearly all the non-patent references (NPRs) on the front page of US patents are for basic research rather than applications, but the citation process can be quite quick –

e-mail: g.lewison@soi.city.ac.uk

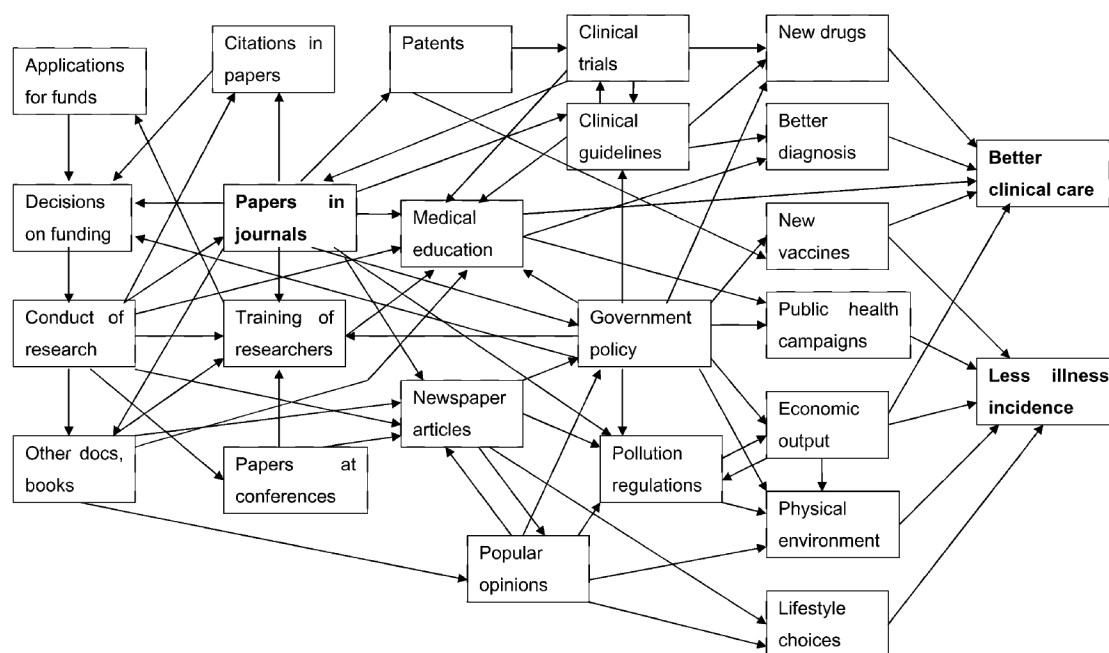


Figure 1. Diagram showing the many linkages between biomedical research and better health.

almost as fast as for citations in journals. The frequency of occurrence of NPRs varies greatly with the subject area, being highest for the applications of medical biotechnology, especially genetics, and very low for mechanical engineering¹². (There is, of course, a corresponding variation in conventional citation scores, but it is not so extreme.)

This paper will discuss five other citation indicators that can be used for research evaluation and give a few results from recent studies to show how comparative values can be calculated. These are very much by way of examples rather than definitive data. The five types of document whose references can be analysed and traced back to research papers are:

- International standards, such as for pesticide residues in food that is to be traded internationally.
- National government policy documents, such as a recent UK Prime Minister's Strategy Unit Paper¹³ on curbing the harm from alcohol abuse.
- Clinical guidelines used to inform medical practice, such as a recent Scottish one on the treatment of alcoholics in a primary care setting¹⁴.
- Textbooks used for the training of medical personnel, which often consist of specialist chapters with references to recent research papers.
- Newspapers, which are increasingly interested to report scientific developments, especially in biomedicine, and which are widely read – by policy makers, by healthcare professionals, by researchers¹⁵ and by the general public.

International standards

This indicator differs from the others in that it is, as its name implies, international. There are several bodies that publish such standards, such as the International Standards Organization (ISO) – though references in them to research are not common – the International Commission on Radiation Protection (ICRP), which is formally a charity but is universally respected for its impartiality, and the Codex Alimentarius, a body set up by the World Health Organization (WHO) and the Food and Agriculture Organization (FAO). Within Europe, the European Union has developed numerous standards and regulations in order to facilitate international trade in many different products. Most of these standards are published on the Web, although for some there is a fee. They can be downloaded and the list of references examined.

Not all the references are to papers in journals, but usually the majority of them are to such research articles. The procedure for their identification begins with the allocation of serial numbers to all the references, and then they are listed on a sheet in date order. They can then be searched for in an electronic database, such as the *SCI*, year by year and their bibliographic details downloaded to file. It may also be useful to look the papers up in libraries (or, if available, as full text on the Web) in order to record details of their funding sources. (This could reveal, for example, the influence of industry on the development of international standards.) The file of cited references must then be compared with an appropriate norm group of papers on a similar subject, produced by

means of a specially-designed filter applied to the chosen database, and consisting of lists of specialist journals and title words. Such a comparison would show if papers from the target group (e.g., a country or research institution) are cited as frequently as would be expected if the standards cited papers randomly from the relevant world literature. Normal statistical tests can then be applied to show whether the observed number of citations to the work of the target group differs significantly from the expected number.

In a study in 2001 on pesticides in food^{16,17}, Atiogbe found that most countries, including India, were represented among the papers cited by the standards about as often as would be expected, but that Spain and France were significantly under-represented, and that the USA was over-represented. This may have been because there was a long delay between the publication of the papers and their citation on the residues standards. At the time when the cited papers were published, the international pattern of research in the subject was rather different, with the USA doing relatively much more, and Spain, much less. This illustrates the importance of being aware of the timing of the citation peak.

Government policy documents

It has become something of a mantra among officials and others concerned with public policy that it should be firmly evidence-based. The opposite, which seems to occur rather too often, is policy-based evidence making, a dangerous practice but one that some politicians prefer. The problem of alcohol abuse is widespread except in predominantly Muslim countries, and the WHO estimate¹⁸ that it accounted for over 10% of Disability-Adjusted Life Years (DALYs) in established market economies in 2000, almost as much as the total attributable to tobacco, which was 11%. In the UK, whereas about 30% of adults smoke, over 90% of adults drink alcohol, so policies to curb binge drinking need careful research if they are to be successful – the simple remedies of steady increases in taxation and more and more restrictions on smoking in most public spaces would not work for alcohol. In March 2004 the Prime Minister's Strategy Unit published an 'Alcohol Harm Reduction Strategy for England', which was based on a carefully-researched Interim Analytical Report, with 215 references, of which 71 were to journal articles and 30 of these (42%) were from the UK. Does this reveal UK alcohol-related research to be sufficiently prominent in helping to inform government policy? The UK is actually poorly represented internationally in such research, with only 6% of the total compared with 10% in biomedicine generally¹⁹. Typically UK documents of the types considered in this paper over-cite UK research by a factor of three or four, so we would expect about 20% of the references to be from the UK, less than half the percentage actually found.

It is also of interest to see what level of research was represented among the cited papers. This can be characterized on the basis of the mean research levels of the journals in which they were published²⁰, or as a cumulative exceedance graph. Figure 2 shows this for UK alcohol-related research and also for all the journal articles cited in the interim analytical report. It is clear that the latter are much more clinical (RL with lower values) than the generality of UK alcohol-related research, and even more clinical than the references on a clinical guideline for alcohol treatment (see below). This suggests that the research that is likely to influence government policy (see Figure 1) will be sociological and clinical rather than basic, but such research is usually rated poorly on conventional citation measures, so it may have difficulty in obtaining funding.

Clinical guidelines

As the cost of healthcare rises, governments and also private insurers are concerned that the treatments provided are both effective and cost-effective. To this end, a number of organizations have been producing clinical guidelines, i.e. recommendations on how to diagnose and treat a variety of medical conditions, based on the evidence from clinical trials and other sources. Such evidence is carefully weighed by those responsible for the guidelines, who may be professional associations (typical in Canada) or governmental bodies. In the UK there are two series of clinical guidelines, one for Scotland produced by the Scottish Intercollegiate Guidelines Network (SIGN) which has issued 82 so far²¹ (although a few have been withdrawn or superseded by others), and one for England and Wales produced by the National Institute for Health and Clinical Excellence (NICE), whose tally²² is currently 91. A report on the guidelines produced by these two bodies up to November 2002 was presented²³ in Beijing in the summer of 2003. It was clear that the papers they cited were predominantly clinical (rather than basic), fairly recent – particularly the NICE ones – and with UK research over-cited by a factor of about three.

Among the SIGN guidelines was no. 74, entitled, 'The management of harmful drinking and alcohol dependence in primary care': this had 158 references of which 88 were in journals and 24, or 27%, were to UK research, which was therefore over-cited by a factor of four relative to the UK's presence in alcohol-related research. Other countries' contributions could be compared with their publications on alcohol-related research in recent years. Figure 3 shows the relative percentages for both the policy document references (IAR) and the SIGN guidelines. Research from the USA is well-cited on these other documents, as is that from Canada and Australia in relation to their world-wide presence, but that from non-Anglophone Member States of the European Union (EU14) and especially from Japan (JP) is relatively neglected.

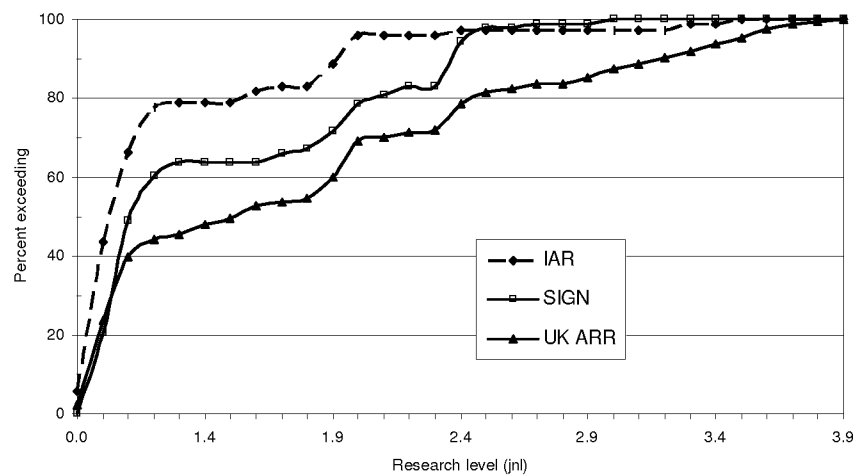


Figure 2. Cumulative exceedance curves for three groups of alcohol-related research papers: UK ARR = all UK alcohol-related research from 1992 to 2003 ($n = 1253$); SIGN = journal references on SIGN guideline concerning treatment for alcoholism ($n = 88$); IAR = journal references on Interim Analytical Report on Alcohol Harm Reduction Strategy for England ($n = 71$). Journal research level determined by calculation based on title words.

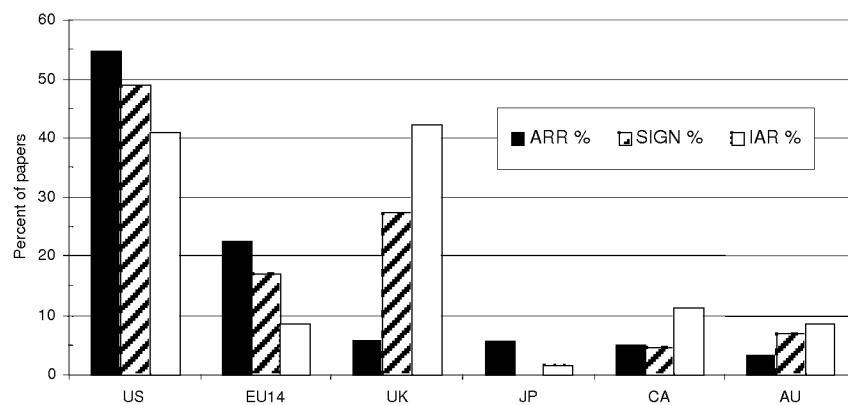


Figure 3. Percentage of papers in alcohol-related research: ARR = world file, 1992–2003; SIGN = journal references on SIGN guideline; IAR = journal references in Interim Analytical Report. EU14 = European Union Member States up to 2004 excluding the UK; JP = Japan, CA = Canada, AU = Australia.

The overall file of references on these UK clinical guidelines can also reveal information on the utility of research *within* a country. Figure 4 shows the percentages of SIGN and NICE guideline citations (up to November 2003) to Swedish research that come from the six medical schools, and the percentages of Swedish health research output in three recent years from the six. It appears that Göteborg and Linköping are performing rather well on this criterion, with about twice as many guideline citations as expected; Lund, Uppsala and Umea have guideline citations in line with their output, but Karolinska, which carries out mainly basic research, is under-cited by about a factor of two. Of course, the more appropriate indicator for this evaluation would be the frequency with which the universities' work was cited on *Swedish* clinical guide-

lines, not British ones; but even the present data show the different medical schools in a new light.

Textbooks

There is now a large range of textbooks, both aimed at the undergraduate level and at more specialized post-graduate studies. They are published in many languages, although in some small countries the demand may not justify the cost of translation from English and other languages. Almost all the references are to journal articles in the more technical subjects, such as schistosomiasis²⁴, though in the 'softer' areas such as the social sciences and gender issues²⁵ this ratio falls, and books become a more impor-

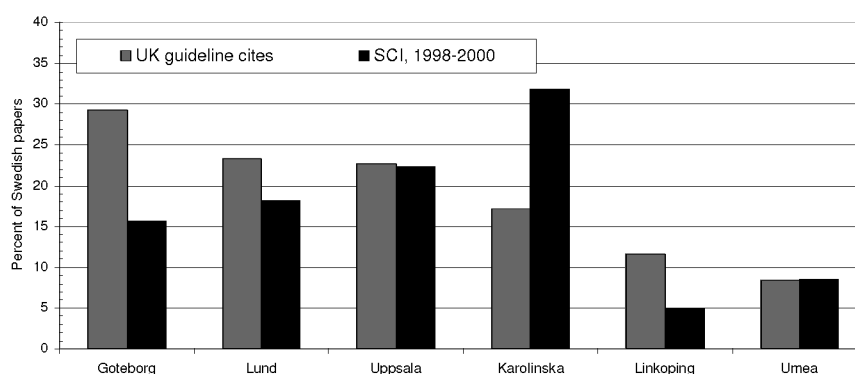


Figure 4. Relative performance of six Swedish medical schools, in terms of numbers of papers cited on NICE and SIGN clinical guidelines (UK guideline cites) and presence in the *Science Citation Index*, 1998–2000: percentage of all Swedish papers ($n = 455$ and $n = 23132$).

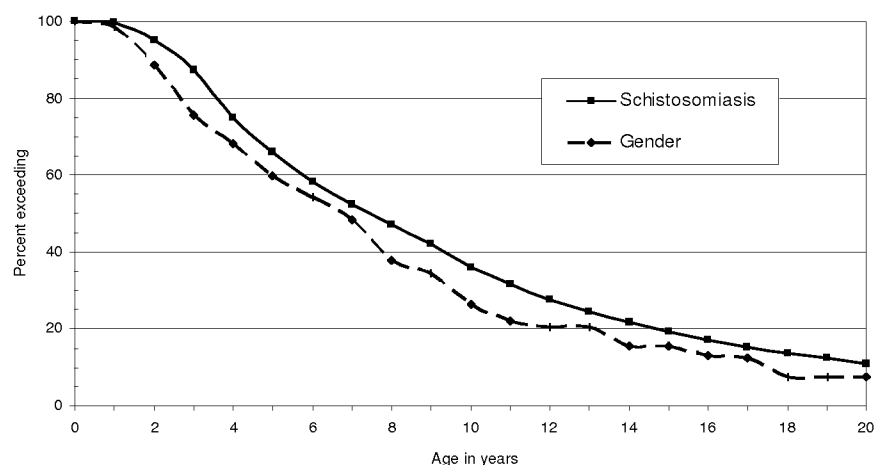


Figure 5. Distributions by age of references in two recent medical textbooks, *Schistosomiasis* (2001) and chapters 1–4 of *Gender Inequalities in Health* (2000).

tant source of information. References tend to be a little older than those on scientific papers, especially on subjects such as tropical medicine which do not receive a lot of current attention from researchers, see Figure 5.

Newspapers

The purists among the bibliometric community may decry the use of newspapers as a means to evaluate research, but they have many attractions as a new indicator of research relevance²⁶. First, news articles partake many of the characteristics of journal articles: they have an author – sometimes more than one – a title, a bibliographic source (newspaper name, date, page(s)), and one or more citations to research items. (These are not given tidily at the end and in a standard format, but there are usually enough clues that the cited article can be identified, sometimes with the help of a news article in another newspaper.) Second, newspapers can be studied at leisure (unlike radio and television, which are ephemeral and

hard to analyse) and many of them are now available for study in full-text versions in commercial databases. Third, it is usually possible to characterize the news article as being primarily based on the research article, or as just mentioning it in passing. So the amount of attention given to the research, coupled with the size and prominence of the news article, can be characterized – something that is not usually possible with citations in journal articles. Fourth, the tone of the news article can be coded – as optimistic, or neutral, or perhaps pessimistic or critical. (In practice, most medical stories in UK newspapers tend to be optimistic, and extrapolate perhaps too far from the small-scale animal model experiments reported to the possibility of a cure for cancer, or some such wonder.) Fifth, the newspaper's circulation and the socio-economic profile of its readership can be obtained, so that an analysis of the different research news presented to different groups of people within a country can be made. And finally, newspapers exist in almost every country, even those with a rather strict censorship, and they do re-

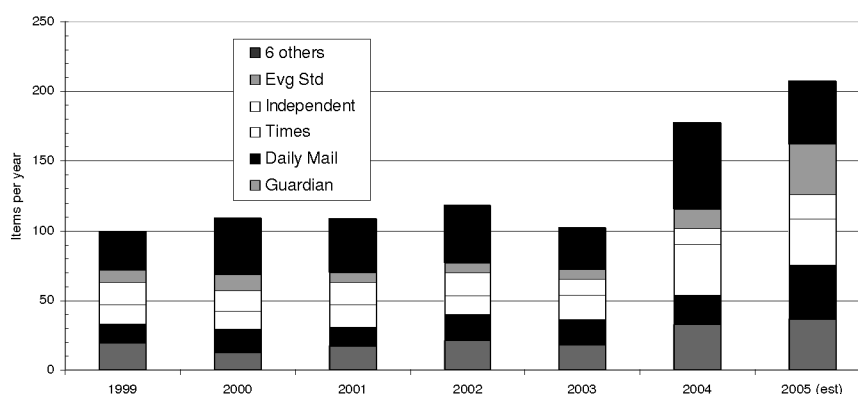


Figure 6. Coverage of alcohol-related research stories in news articles in 11 British newspapers, 1999–April 2005. Total for 2005 has been extrapolated from numbers in first four months of the year.

port biomedical research because biomedical research is almost universally regarded as a public good, and something to be encouraged, even if not very much takes place within a given country.

There is, as one might expect, quite an extensive literature on how newspapers report science, and biomedical research in particular, but almost all such studies have been of specific topics, and have been for relatively short periods of time. An exception was the study by Adelman and Verbrugge²⁷ which looked at how US newspapers covered six major diseases over a 20-year period; they showed pretty conclusively that ‘death makes news’ and the three relatively non-lethal diseases (diabetes, arthritis and Alzheimer’s) received much less attention than the three lethal ones (heart disease, AIDS and cancer). Outbreaks of unusually deadly diseases seem to be particularly newsworthy, but there is less reporting of relevant research – perhaps because there is not much of it. There is also a clear tendency for newspapers to try to present a scientific controversy as being a battle between roughly equal opposing forces, even if the scientific consensus clearly favours only one side. Examples are the recent controversy in the UK over the mumps, measles and rubella (MMR) vaccine^{28,29} and, in South Africa, over whether HIV is responsible for AIDS^{30,31}.

Newspapers are particularly interested in social problems, especially if they affect us all, so the recent debate in the UK over ‘binge drinking’ (mainly by young people) and the possibility of being able to do this 24 hours per day has attracted a lot of press coverage. We analysed coverage in 11 UK national newspapers (including two important regional ones) of alcohol-related research during the last six years, see Figure 6. There were a total of 989 news articles from January 1999 to April 2005, and the number was fairly constant at just over 100 per year until 2004 when it jumped up sharply, reflecting public concern over possible 24-hour drinking; this increase in reportage continued into 2005. Among the newspapers, *The Guardian* (a left-of-centre broadsheet) and *Daily Mail*

and *The Times* (right-of-centre tabloids) gave the subject the most coverage, showing that the subject was thought to appeal to many types of reader. More detailed analysis of the news articles would allow the underlying research publications to be identified, looked up in libraries, and classified: this would show the extent to which UK alcohol-related research was being brought to national attention.

Conclusions

We have described some new forms of citation analysis, and how they might be used in the future to provide data that can complement (but not replace) both conventional citation analysis and traditional peer-review. They are not particularly difficult to develop, but it would seem worthwhile to set up one or more national evaluation units that would generate the appropriate methodology, or apply that developed elsewhere, for the production of new research indicators, and the exchange of data with similar units in other countries. Once the indicators have been produced, it will be possible to generate ‘kite’ diagrams showing national or institutional performance on a number of different scales. However the scales to be used for particular types of research will need to be chosen with care in order to reflect the purposes of the research and its typology.

1. Garfield, E., A new dimension in documentation through association of ideas. *Science*, 1955, **122**, 108–111.
2. Lederberg, J., How the science citation index got started. In *The Web of Knowledge – A Festschrift in Honor of Eugene Garfield* (eds Cronin, B. and Atkins, H. B.), Information Today, Inc., Medford, NJ, 2000, pp. 25–64.
3. Leydesdorff, L., Theories of citation. *Scientometrics*, 1998, **43**, 5–25.
4. Hanney, S., Frame, I., Grant, J., Green, P., Buxton, M., Young, T. and Lewison, G., Using categorisations of citations when assessing the outcomes from health research. *Scientometrics*, 2005, **65**(3), 357–379.
5. Benning, S. P. and Speer, S. C., Incorrect citations – a comparison of library literature with medical literature. *Bull. Med. Libr. Assn.*, 1993, **81**, 56–58.

6. Van Raan, A. F. J., Advanced bibliometric methods as quantitative core of peer-review based evaluation and foresight exercises. *Scientometrics*, 1996, **36**, 397–420.
7. Lewison, G., Cottrell, R. and Dixon, D., Bibliometric indicators to assist the peer review process in grant decisions. *Res. Evaluation*, 1999, **8**, 47–52.
8. Martin, B., The use of multiple indicators in the assessment of basic research. *Scientometrics*, 1996, **36**, 343–362.
9. Lewison, G. and Carding, P., Evaluating UK research in speech and language therapy. *Int. J. Lang. Comm. Dis.*, 2003, **38**, 65–84.
10. Narin, F., Patent bibliometrics. *Scientometrics*, 1994, **30**, 147–155.
11. Harhoff, D., Narin, F., Scherer, F. M. and Vopel, K., Citation frequency and the value of patented inventions. *Rev. Econ. Stats.*, 1999, **81**, 511–515.
12. Narin, F. and Olivastro, D., Linkage between patents and papers: an interim EPO/US comparison. *Scientometrics*, 1998, **41**, 51–59.
13. The Prime Minister's Strategy Unit, Alcohol misuse: How much does it cost? Annexe to the Strategy Unit Alcohol Harm Reduction project Interim Analytical Report. The Cabinet Office, London, 2003.
14. <http://www.sign.ac.uk/guidelines/fulltext/74/index.html>.
15. Phillips, D. P., Kanter, E. J., Bednarczyk, B. and Tastad, P. L., Importance of the lay press in the transmission of medical knowledge to the scientific community. *N. Engl. J. Med.*, 1991, **325**, 1180–1183.
16. Atiogbe, P., Bibliometric analysis of the international Codex Alimentarius food pesticide standards. M Sc dissertation, Department of Information Science, City University, London, 2001.
17. Lewison, G., Beyond outputs: new measures of biomedical research impact. *Aslib Proc.*, 2003, **55**, 32–42. See pp. 34–37.
18. World Health Report 2002: Reducing Risks, Promoting Healthy Life. Geneva, WHO.
19. Rajendram, R., Lewison, G. and Preedy, V., Worldwide alcohol-related research. *Alcohol and Alcoholism*, in press.
20. Lewison, G. and Paraje, G., The classification of biomedical journals by research level. *Scientometrics*, 2004, **60**, 145–157.
21. <http://www.sign.ac.uk/guidelines/published/numlist.html>
22. <http://www.nice.org.uk/page.aspx?o=cg>
23. Lewison, G. and Wilcox-Jay, K., Getting biomedical research into practice: the citations from UK clinical guidelines. In Proceedings of the 9th International Conference on Scientometrics and Informetrics, Beijing, 2003, pp. 152–160.
24. Mahmoud, A. A. F. (ed.), *Schistosomiasis*, Imperial College Press, London, 2001.
25. Annandale, E. and Hunt, K. (eds), *Gender Inequalities in Health*, Open University Press, Buckingham, 2000.
26. Lewison, G., From biomedical research to health improvement. *Scientometrics*, 2002, **54**, 229–240.
27. Adelman, R. C. and Verbrugge, L. M., Death makes news: the social impact of disease on newspaper coverage. *J. Health Soc. Behav.*, 2000, **41**, 347–367.
28. Hargreaves, I., Lewis, J. and Spears, T., *Towards a Better Map: Science the Public and the Media*. UK Economic and Social Research Council, Swindon, 2003. Available online at: <http://www.esrc.ac.uk/esrccontent/DownloadDocs/Mapdocfinal.pdf>
29. Brown, H., MMR drama outrages UK doctors. *The Lancet*, 2003, **362**, 1986.
30. Cohen, J., South Africa: AIDS researchers decry Mbeki's views on HIV. *Science*, 2000, **288**, 590–591.
31. Sidley, P., Mbeki plays down AIDS and orders a rethink on spending. *BMJ*, 2001, **323**, 650.

Quotations by Eugene Garfield

'One of the lessons I learned is that there is enormous resistance in both commercial and academic organizations to change, especially if there are entrenched empires – even if they are more expensive services. It took about ten years before the last big holdout, Merck, converted to Current Contents. They bought 500 copies for at least two decades. Today they are still an important customer'.

— Garfield

On The Origins of Current Contents and ISI Presented at the 12th Annual Meeting of the Association of Independent Information Professionals, St. Louis, Missouri, 1998

'I have always stressed that Citation Indexes are no substitute for subject indexes. This is true of the legal literature too. First one uses the "digest" to find an interesting case of two and then uses the citator to locate the cases that have subsequently emanated from them.'

— Garfield to Joshua Lederberg

In *The Web of Knowledge – A Festschrift in Honour of Eugene Garfield*
Information Today Inc, Medford, 2000

'I can't understand people – why would a person spend a fortune on lab equipment and not \$50 on a service that saves him his most precious possession – time.'

— Garfield to Joshua Lederberg

6 July 1959