Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 54, No. 2 (2002) 179–192

From biomedical research to health improvement

GRANT LEWISON

Bibliometrics Research Group, Department of Information Science, City University, London EC1V 0HB (England)

Traditional means of analysis of research outputs have focussed on citations to papers in journals in other journal publications. But these only chronicle the early stages whereby research in biomedicine is converted into health improvement through better patient care and through preventive measures. New evaluation methods, still based on the concept of citation of research in other documents, are needed and are now being developed. These include the use of textbooks in medical education and the analysis of governmental regulations and health policies, which can influence both the availability of new drugs and the control of toxic substances in food and the environment. There is also an interest in the way that newspapers report biomedical research advances. Readers include politicians, healthcare professionals, the general public (who are increasingly becoming active consumers of healthcare products) and other researchers who may value the immediacy of the reporting. Newspaper articles tend to focus on fashionable topics and to offer premature hopes of cures to disease, but they can also provide a valuable service in showing the importance of animal experiments to biomedical progress. It would be useful to create an international database of newspaper citations through a consortium of partners in different countries who would agree a common protocol and exchange information on a regular basis.

Introduction

Bibliometricians have mainly confined their attention to the publications that result directly from research and their various properties, including especially their references to earlier work. The subject has been made possible largely as a result of the availability of databases of scientific articles in journals and in particular of the *Science Citation Index*, which Eugene Garfield set up more than three decades ago and is still regarded as the "gold standard" for scientific publications. However although many bibliometricians remain fascinated by the *minutiae* of citation analysis, its uses and its limitations, those who are concerned with science policy need indicators that relate more to the real world of innovation and healthcare.

I thought that it would be useful for us to look up from our detailed studies and try to survey the wider landscape, at least in one area with which I am familiar, namely biomedical research and its relationship to improved health. The limitation to this topic

0138–9130/2002/US $ 15.00

is not so restrictive as it might appear as much commercial innovation depends on development work carried out by companies in conditions of secrecy, where there is no public record of the research that has taken place. By contrast in biomedicine the practice of publishing one's results in the open literature prevails not only in academia but is also encouraged in the commercial world. Large pharmaceutical companies need to attract excellent research staff who wish to develop their individual scientific reputations and careers. Small biotech firms want to advertise their technical prowess to potential partners. So, with some exceptions, the serial literature gives a good impression of the volume and scope of biomedical research that has taken place.

But there is a long and tortuous route from papers in journals, particularly in ones concerned with basic research, to the improvement of patient care and the prevention of disease, the twin goals of biomedical research. Figure 1 shows some of the steps on the way and the linkages between them. It is inevitably a much simplified diagram and we could add many more steps and linkages but then it would be hard to decipher.

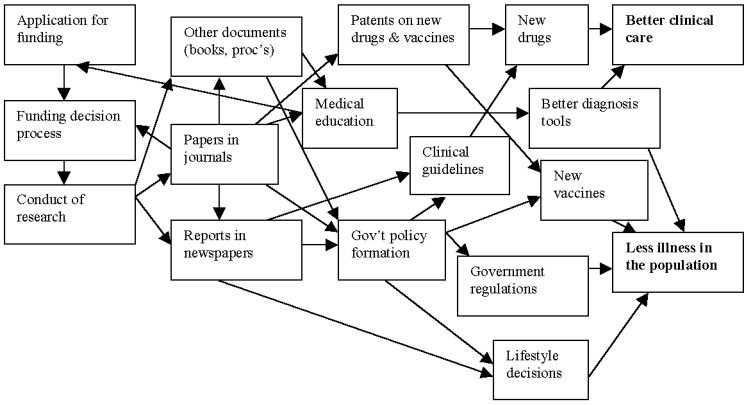

Figure 1. The links between biomedical research and better health

The final goals are on the right side of the diagram. Currently, much biomedical research is directed towards the upper goal, better clinical care, and it seems to have a more attractive image, with doctors in white coats or surgeons in green robes intervening heroically to save a patient from some dreaded disease. But historically far more lives have been saved by simple preventive measures, of which the removal of the Snow Hill pump handle to prevent the spread of cholera in 1854 London is perhaps the

best-known example. The increasing number of restrictions on tobacco advertising and use in public may have a similar effect in health improvement, albeit on a long-term basis.

There are, however, some high-tech preventive measures which depend on high quality biomedical research. A recent UK example is the rapid introduction[1,2] in November 1999 of the vaccine for meningitis C, which was originally developed by the Department of Health. When the vaccine became available, it required a major political effort to secure funding to immunise 15 million children within a year, and a corresponding activity in publicising the vaccine's availability and distributing it to many thousands of nurses. The effect on the incidence of, and deaths from, this disease among children was dramatic.

Although there have been many studies that link research to innovation, such as through the citation of research papers on patents for new drugs, and there is a lively discussion on whether the "linear model" is still relevant, this alternative pathway to better health is much less trodden. I would like to suggest that bibliometricians should attempt to follow it and provide maps of the route, with indicators of progress that can be used to evaluate the utility of research for health improvement in this way.

Citations still rule OK

This does not mean that the traditional techniques of the bibliometrician, counting publications and citations, are no longer appropriate. It just means that she or he will have to work with different media, and collect information directly rather than relying on established databases that can be searched electronically. Links between documents can still be determined through the principle of referencing; they may need to be supplemented by social surveys to establish which documents were instrumental in informing a decision. Thus a political decision to change pollution control legislation, or to amend food standards, or to tax leaded petrol more highly than unleaded fuel, or to phase out X-ray machines in shoe-shops, or to make seatbelts in cars compulsory, will need to be investigated in detail to learn what evidence was truly influential.

This will give new meaning to "citation analysis" and it should ensure a bright future for bibliometricians who can adapt to the different environment for their work. This morning I should like to discuss three projects that we are currently carrying out at City University to develop new ways of working and to think about the new questions that can be asked and answered. Two of them are at a very preliminary stage but should yield some sample results later in the year. One of them is also preliminary but has started to provide some results and may stimulate other bibliometricians to copy it in

their own countries. This project is essentially national but the methodology could be replicated elsewhere and the resulting database would thereby increase in value. The other projects are on a small scale and again could be added to in other centres in order to allow a more comprehensive view of this particular part of the citation linkage diagram.

The point I am making is that, whereas most databases used by bibliometricians are international in scope and are disseminated by a single organisation, either commercially or *pro bono publico*, these new databases will be essentially created in many centres. There is widespread use of Microsoft database and spreadsheet programs and universal access to the world-wide web and the internet. This means that it is now possible for bibliometricians in different countries each to work on a small part of a database, pool their results and have access to the work of many others for their analysis.

The common thread in all this work, however, will be the concept of citation, with a document citing one or more source documents. However the identification of the source documents may not be trivial: the documents we shall be dealing with may not use the Vancouver convention or even give a clear and unambiguous reference that can be matched to a paper in a peer-reviewed journal. We shall need to track down these sources so that their use in later documents can be chronicled and credited to them. Conventional bibliographic databases may help in this task but sometimes detective work, aided by a good memory, may be equally useful.

Government food regulations: the Codex Alimentarius Commission (CAC)

This organisation was set up in 1962 by the World Health Organization (WHO) and the Food and Agriculture Organization (FAO) and it includes almost all United Nations members. Its work includes the publication of many standards that are used for the international trade in foods. We are looking at one of these, the standard covering pesticide residues. There are three main parts, each supported by lists of technical reports and papers. They are:

- methods of analysis for the determination of various pesticides in foods;
- toxicology on the maximum quantities of pesticides that can be safely consumed; and
- dietary surveys to establish the likely patterns of consumption of particular foods.

Although this work may seem bureaucratic and remote, performed as it is by experts sitting in apparently tedious and interminable committees, it underpins the safety of

many of the foods we routinely eat. When we purchase beans grown in Kenya or coffee from Colombia in our local supermarket we can be reasonably sure that they will be fit for consumption. This is an important safeguard: food contamination, whether bacterial, chemical or by heavy metals, has often been a serious public health risk in the past as witness the cooking oil scandal in Spain and the mercury-contaminated fish in Minamata, Japan.

Our project will look at the pattern of the references and seek to determine how old they are, from which countries the research emanates, what type of research (which journals) is cited and perhaps, why the committee chose those particular papers to cite. Naturally, this will be only a small sample of the total output of the CAC, and it will in due course need to be extended to cover all the regulations and supplemented by examination of national food regulations and their evidence base. However we hope to establish a convenient methodology for the compiling of references and for answering questions such as:

- Does the underlying science depend on work funded by government or by industry?
- Is the science published in journals of high repute?
- How long does it take for food safety work to be accepted and cited in regulations?
- Does a country gain commercial advantage from having carried out such research?

Once the methodology has been established, we shall invite other interested organisations to join with us to add their data to a common database.

Medical education: the use of textbooks

Although books have long been regarded as an important output of research, bibliometricians have found them difficult to process. This is partly because the references to and within them are so scattered but also because their value as records of biomedical research is rather indirect. Yet they continue to be published, usually nowadays in the form of collections of individual essays or papers presented at conferences rather than monographs.

We intend to explore how an individual textbook, *Autonomic Failure*, edited by Bannister and Mathias (Oxford University Press, 4th edition, 1999) cites the underlying scientific literature in its 19 chapters. But equally, we intend to find out what use is made of the book (and its preceding editions) in medical education and in continuing clinical practice through questionnaires to physicians and researchers and some interviews. This is, of course, a specialist book, but one widely seen as authoritative in its sub-field (autonomic nervous system) and therefore potentially a good guide to

doctors in their clinical practice. If the project succeeds in tracing the links between research and practice, we shall be able to answer questions such as:

- Is clinical practice informed by recent research? If so, does the textbook play a big part in this?
- Which countries' research is of particular use in medical education and does the selection reflect the overall balance of output in the subject?
- To what extent does autonomic nervous system education depend on research in the specialty, or on work in other areas?
- Is the research cited clinical or basic and are there differences between the times needed for the different types of work to be cited in the textbook?
- How can recent research be best communicated to practising clinicians?

Again, once the methodology has been established, we would like to extend it to other books including undergraduate texts, and invite bibliometricians in other countries to work in similar ways so that we can see whether medical training differs between countries. Since within the European Union, medical qualifications are now transferable between countries, this work will have an international dimension. It will allow us to see if doctors trained in another Member State have been taught to depend on a different body of evidence when making their clinical judgments: this might be salutary or it could be dangerous.

Newspapers: bringing in the public

Although much attention is focussed on the internet as a source of information, its penetration is still far below that of the mass media such as television, radio, newspapers and magazines. The internet provides a huge mass of information, and there are now special sections of bibliometrics devoted to its analysis. However it suffers from one major defect in that it is constantly changing and experiments are seldom repeatable as a result. It is also unclear how one can analyse the "hits" made on a given web-site: can they be regarded in the same way as requests for a document in a library, for example? And since some of the information placed on the WWW is of very dubious quality, it is hard to say that real information has been transferred when someone has downloaded information from a web-page, let alone if they have merely accessed it.

Television and radio have the virtue of being rather more definitive and they have the further advantage that their audiences are well researched so that the numbers of people watching or listening to a particular programme can be estimated quite well. The difficulties for the bibliometrician are in monitoring and in deciding what a citation is.

There are some programmes where recent science is presented and discussed, and where one can trace the source to a particular scientific paper, or more likely to several related papers. But these are likely to have much smaller audiences than the general news and comment programmes in which scientific discoveries are treated as news, and a scientific journalist is invited to comment on the significance of the results or interviews the researcher involved. For example, the British Broadcasting Corporation's Radio 4 programme "Today", which is broadcast on weekday mornings, often has reports based on a recent paper in *Nature*, *Science*, the *BMJ* or *Lancet* which appear on Fridays. These reports will reach a very wide audience and may therefore attract a lot of attention. They can be traced quite easily to the original article. However, the job of monitoring all the output of the main radio and television channels for their references to science, identifying the sources and then classifying the citations in some way is rather daunting.

We have decided instead to focus attention on national newspapers. These have many advantages for the bibliometrician:

- they are cheap and readily available and normally change little between editions;
- they are widely read and their readership is well characterised both socially and geographically. It includes doctors, health policy people, politicians and the general public who are increasingly treating clinical care as a consumer service to be selected on the basis of good information;
- the citing articles have many of the characteristics of a scientific paper, namely they have an author (occasionally several, sometimes none), a bibliographic description (source), title and length;
- the cited sources are usually identified reasonably clearly, although some references are a bit oblique or merely *en passant*.

In the UK there are some 20 national newspapers, including ones published on Sundays which are separate from those appearing Monday to Saturday. They are regularly monitored by the Wellcome Trust's Information Service in order to inform Trust employees on current medical stories in a daily collation called *What the papers say* (WTPS). This reproduces articles on a number of subjects but always includes reports on current research that appear in the press. Because some stories appear in many newspapers, only a selection is made for WTPS but the others are also clipped and can be recorded in a database.

This is now being maintained, initially in the form of an Excel spreadsheet, and some results have been obtained for the two months, April and May 2001. They may not be fully representative, but they give an idea of the characteristics of the reports and, more important, they indicate the questions that can fruitfully be asked. Some of the

findings are given below and they suggest all sorts of new lines of enquiry that could be the subject of research by our group, and others, in the future.

The basic record consists of a newspaper article citing a research finding. One article may cite several findings (i.e., several separate research publications) but most cite only one and are mostly based on this one paper. Often the journalist will extrapolate from research on animals to suggest that a new treatment for a disease is possible, and may engender false hopes – typically of a "cure for cancer". Many articles contain human interest stories of patients who have been treated with an experimental drug and derived benefit from it; these are likely to stimulate readers to demand similar treatment from their general practitioner. Some articles contain cautionary notes from a spokesman from the relevant disease-related charity warning that it will be many years before the treatment described is routinely available. In the meanwhile, of course, it is vital that more research is carried out and that people continue to give to the charity.

Each record includes the following data:

- serial number;
- date of newspaper article;
- newspaper (coded with a digraph, e.g., DT = Daily Telegraph, GU = Guardian)
- page number;
- article length in column cm. When the article refers to several pieces of research, the length of the part of the article referring to each is given;
- journalist name, written as it appears e.g., James Chapman;
- title of article, without single quotes which often appear. Sometimes a sub-title is added if the article is long and the title is short but the sub-title is more informative;
- description, of what the original article said, in about 20-40 words. This should indicate if the experiments described were on animals and if so on which species and give the names of drugs as used in the literature (e.g., Fluoxetine rather than Prozac);
- whether the citing article gave background information on the disease and its incidence, on a scale from 0 to 4;
- whether it gave details of the methodology used in the research, again on a scale of 0 to 4;
- whether it gave some evaluation of the likely limitations of the research results, on the same scale;
- the names of any independent commentators cited who could provide a context for the research findings, their affiliation(s) and whether they were supportive or critical;
- research author(s) cited, in SCI format, e.g., Watson-J Crick-A;

- sub-field(s) of interest. About 60 are used which correspond largely to medical specialties or biomedical areas; they are written as pentagraphs in upper case, separated by spaces e.g., GENET ONCOL;
- journal cited, in SCI 29character format, e.g., BRIT MED J, N ENGL J MED;
- addresses of author(s), in SCI format, e.g., HARVARD-UNIV, DEPT PSYCHOL, CAMBRIDGE, MA, USA. Often only sketchy details will be available, with the name of the university but not the city;
- country/ies of the author(s), as ISO digraphs separated by spaces, e.g., CA DE UK;
- year, normally the current year but sometimes earlier;
- notes. Sometimes the paper will be a contribution to a conference but will appear later as a journal article which will form the definitive record. There may be details given of which issue of the journal contains the cited paper;
- funding. This is sometimes given in the citing article; it may need to be supplemented by inspection of the original source.

All this information can be recorded from the newspaper article though often some of the fields will be blank. Further information can be added to enrich each record:

- from the cited source document: all authors' names, title, bibliographic citation, all addresses, funding acknowledgements;
- whether the paper was subject to one or more Press Notices. These might have been issued by the journal, by the funding source for the research, or by the lab of the researchers. The similarity of many of the press articles suggests a common source for the journalists' information;
- any feedback obtained by the journalists as a result of their articles. Very occasionally there will be correspondence in the newspaper following publication of the article but some journalists give an e-mail address which seems to invite correspondence. Feedback can also go to disease-specific charities, which are often inundated with telephone calls following an optimistic newspaper article.

Sample results from newspapers

In this final section, I should like to give some of the preliminary results from the analysis of two months' press articles in the UK. They may serve to show what others should look for if they decide to adopt a similar approach to the recording of information in the national press. During this period (61 days) there were 396 citations (6.5 per day); they were very unequally distributed between papers as Table 1 makes clear. This table also shows the mean article length in column-cm.: somewhat

surprisingly, there is relatively little variation in this. Figure 2 shows a histogram of article lengths. The mean is 38 column-cm and the median is 33 column-cm. In total, 137 article-cites were in tabloids and 259 in broadsheets.

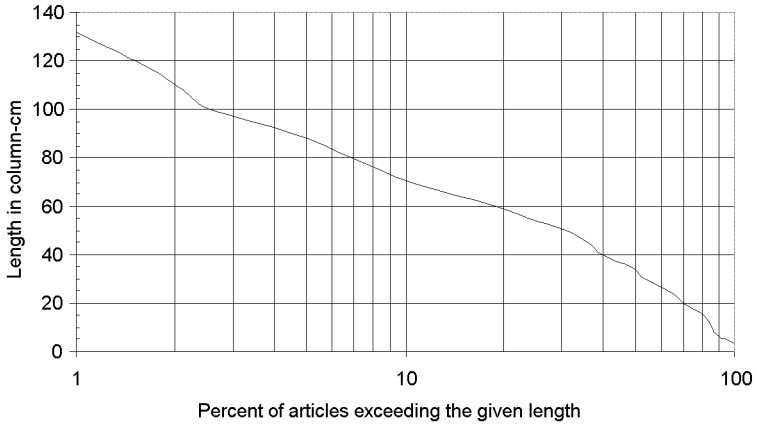

Figure 2. Lengths of articles in UK national papers reporting biomedical research

Table 1. Reports of biomedical research in UK national papers, April-May 2001: numbers of articles per day and mean length in column-cm.

Code	Newspaper	Art/day	Mean L	Code	Newspaper	Art/day	Mean L
DM	*Daily Mail*	1.54	44	SC	*The Scotsman*	0.54	43
DT	*Daily Telegraph*	1.33	32	SG	*Sunday Telegraph*	0.44	41
TI	*The Times*	1.04	33	SU	*The Sun*	0.44	38
ST	*Sunday Times*	1.00	24	MS	*Mail on Sunday*	0.33	39
OB	*The Observer*	0.89	56	FT	*Financial Times*	0.19	23
IN	*Independent*	0.87	49	MI	*Daily Mirror*	0.13	30
DE	*Daily Express*	0.69	38	IS	*Independent on Sunday*	0.11	66
SM	*Sunday Mirror*	0.67	19	SE	*Sunday Express*	0.11	60
GU	*The Guardian*	0.60	35				

One of the initial surprises was the large number of individual journalists who reported on medical research. Of the 396 article-cites, 86% had a named author and there were no fewer than 110 different writers but only 10 of them wrote at least one article per week on average. Among the 55 different sub-fields identified, with a mean

of 1.6 per citation, the leaders were dietetics (18% of citations), oncology (17%), neurosciences (16%), epidemiology (11%), genetics and obstetrics & gynaecology (10% each), and pharmacy & pharmacology (8%). Cardiology, despite being a major killer, only accounted for 6% of citations, little more than mental health and infectious diseases (5% each).

An analysis of the words in the titles revealed the pattern shown in Table 2. (Only 18 article-cites had no title.) The general tone of the articles was optimistic, with 27 mentions of hope/hopes and 19 of good/better compared with only 3 of bad/worse. The articles tended to be aimed at women with 19 mentions compared with only 5 for men; 12 for breast but only 1 for prostate. They were also concerned with the young with 23 mentions of children and 19 of baby/babies but only 1 of old. They tended to be personalised and had 32 mentions of you/your/you're compared with only 11 of they/their/they're.

Table 2. Most frequently-used words in titles of 396 UK newspaper citations of biomedical research

Word(s)	N	Word(s)	N	Word(s)	N
Cancer/s	51	Risk/s	21	New	18
You/your/you're	32	Baby/ies	19	Scientist/s	16
Hope/s	27	Drug/s	19	Test/s	16
Gene/s/genetic	25	Good/better	19	Heart	14
Child/ren/hood/'s	23	Woman/women	19	Breast	12

A common theme was to extrapolate results from animals to humans. Of the 396 article-cites, 8% involved animal experiments, mainly mice and rats. The general tone was very supportive of the need for and utility of rodent experiments as models of human physiology and disease patterns: this may be important in securing continuing public support for such work in the face of violent protests by animal rights people.

Locating the sources of the science was sometimes not easy. Typically the country/ies of the researchers were given (see Table 3); only 6% didn't mention any country. There is naturally a strong UK bias, but New Zealand (3 papers, 6 cites) did surprisingly well and France and Japan very poorly.

Next most frequent was a mention of the researchers' address (89%), followed by the name of the team leader (77%) and the journal (55%). However quite a lot of the reports were of papers to conferences which had probably not yet been published in a journal. Not surprisingly, since these were UK newspapers, UK journals predominated in numbers of mentions and occupied the first 5 places in the list (Table 4).

Table 3. Countries of researchers mentioned in UK newspaper articles, compared with their presence in the world biomedical literature (SCI, 2000)

Country	N	%	BM %	Ratio	Country	N	%	BM %	Ratio
UK	168	42.4	10.3	4.1	Italy	6	1.5	4.8	0.3
USA	144	36.4	39.3	0.9	Netherlands	6	1.5	3.3	0.5
Germany	17	4.3	8.6	0.5	New Zealand	6	1.5	0.4	3.4
Australia	12	3.0	2.8	1.1	Switzerland	5	1.3	2.2	0.6
Canada	10	2.5	4.7	0.5	Finland	5	1.3	1.3	1.0
Sweden	7	1.8	2.7	0.7	Japan	5	1.3	10.4	0.1
Spain	6	1.5	2.7	0.6	France	4	1.0	6.3	0.2

Table 4. Journals most frequently mentioned in UK newspaper articles reporting biomedical research. UK publications shown in **bold**

Journal	N	Journal	N	Journal	N	Journal	N
New Scientist	29	***J Epid Com H***	10	*Circulation*	6	***Hum Reprod***	4
Brit Med J	25	*JAMA*	9	***Gut***	6	***J Roy Soc Med***	4
Lancet	14	*N Engl J Med*	9	***Nat Neurosci***	6	***Nat Genet***	4
Nature	11	*Science*	7	*P Nat Ac Sc US*	5	***Psychologist***	4

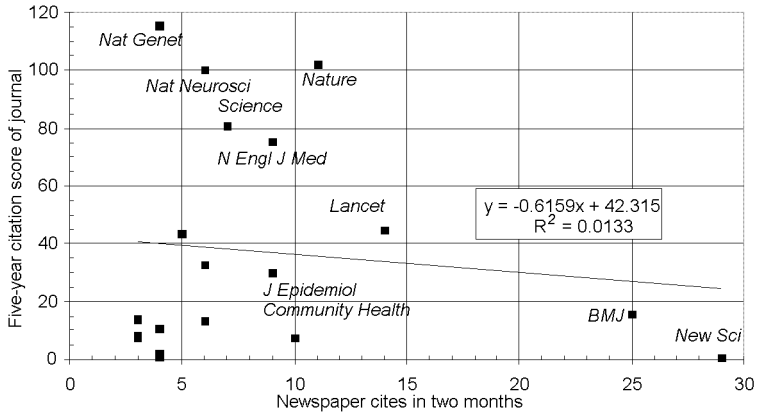

Figure 3. Comparison between citation scores and frequency of citation by name in UK newspaper articles of 16 journals

Figure 3 shows a comparison between the frequency of journals being cited by name in the newspaper articles (some cited articles were in journals but the journals were not named) and their five-year citation scores, i.e., the mean numbers of cites from 1994-98 of papers published in 1994. (These data are taken from the Journal Expected Citation Rates file produced by the Institute for Scientific Information for 1998, the latest year to which I have access.) It is clear that there is virtually no correlation, largely because of the dominance of *New Scientist*, a journal with few conventional citations but one that is designed for lay people and whose articles are therefore easily accessible to journalists.

The last piece of analysis concerns the research level of the journals cited. For the 94 article-cites involving a UK scientist where the journal was identifiable, the division between journals on the basis of the research levels assigned to them by CHI Research Inc. was as shown in Table 5. (Four papers were in a journal, *Psychologist*, that has not been classified.)

Table 5. Distribution of journals cited in UK newspaper reports by research level (RL) and comparison with data for UK research in 1998.

Research level	RL	News cites	UK research	Ratio
Clinical observation	1	40	4255	0.94%
Clinical mix	2	23	6648	0.35%
Clinical investigation	3	5	6258	0.08%
Basic research	4	22	8776	0.25%

There appears to be a preponderance of clinical research cited: papers in journals classed as RL 1 or 2 make up two thirds of citations in newspapers but only 41% of UK biomedical output. Nevertheless, basic research is not entirely neglected, even by the tabloids: of the 84 papers in basic research journals overall, 32% were cited in tabloids, whereas of the 129 papers in clinical journals (RL = 1 or 2), only 26% were cited there.

Discussion

In some ways bibliometrics is at the stage of European navigation in the middle ages. The familiar territory is well, even obsessively, charted but beyond the known world there are only unknown dragons on the map. It is time to use the tools that we have developed to venture forth and discover new continents where we may find great riches. It is thus entirely appropriate that the ISSI meetings are now setting forth into

new locations away from the main scientific trade route of North America, Europe and the middle East. In 1999 we met in Mexico, we are now in Australia and in 2003 we shall meet in China.

The new voyages of exploration will no doubt occasion casualties and blind alleys just as the Renaissance sailors met with heavy losses on their journeys into uncharted waters. But there are now quite a lot of us able to launch expeditions and we can work together to map the new landscape more quickly and thoroughly. I think that we shall find many new indicators of scientific utility and that in the process we shall learn more about our subject and the process by which science gets turned into better health.

*

The author is a Senior Policy Adviser at the Wellcome Trust, now working on secondment at City University. The work on newspaper citations was made possible by the compilation, *What the papers say*, produced by Elizabeth *Graham* and her colleagues in the Wellcome Trust Information Centre, and the author is most grateful to them for their dedicated and careful work. The work on the references in CAC documents is being carried out by Potenza *Atiogbe* and that on the references in *Autonomic Failure* by Helen *Kershaw*, both students for the MSc degree at City University.

References

1. J. BRADBURY, New meningitis-C vaccine to be used in the UK. *Lancet*, 384 (1999) : 310.
2. J. WISE, UK introduces new meningitis-C vaccine. *BMJ*, 319 (1999) : 278.

Received October 12, 2001.

Address for correspondence:
GRANT LEWISON
Bibliometrics Research Group, Department of Information Science
City University
Northampton Square, London EC1V 0HB, England
E-mail: g.lewison@soi.city.ac.uk