Jointly published by Elsevier Science Ltd, Oxford and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 45, No. 1 (1999) 117–136

A COMPENDIUM OF ISSUES FOR CITATION ANALYSIS

T. J. PHELAN

Research Evaluation and Policy Project, The Australian National University, Canberra (Australia)

(Received January 6, 1999)

This paper examines a number of the criticisms that citation analysis has been subjected to over the years. It is argued that many of these criticisms have been based on only limited examinations of data in particular contexts and it remains unclear how broadly applicable these problems are to research conducted at different levels of analysis, in specific fields, and among various national data sets. Relevant evidence is provided from analysis of Australian and international data. Citation analysis is likely to be most reliable when data is aggregated and at the highly-cited end of the distribution. It is possible to make valid inferences about individual cases, although considerable caution should be used. Bibliometric measures should be viewed as a useful supplement to other research evaluation measures rather than as a replacement.

Introduction

Over the years, numerous problems associated with bibliometric methods have been raised in the literature. While discussion of these concerns has been thoughtful and extensive, many important issues have yet to be considered within a sufficiently large number of applied research settings to enable us to know just how broadly applicable these seemingly fundamental problems really are. Unfortunately, there are still relatively few studies specifically aimed at the replication and validation of basic methodological findings across fields, at various levels of analysis, and within different nations. Furthermore, as time passes, old warnings inevitably tend to be ignored or forgotten and new researchers entering the field may not realise that many of these earlier concerns have never been completely resolved.

The focus of this study is on fundamental methodological concerns in the field of bibliometrics. It is intended to provide two useful contributions. First, it will review a number of longstanding methodological issues, and provide new evidence relating to them, from a variety of perspectives, examining both Australian and international data

0138–9130/99US $ 15.00

sets. Second, it will provide within a single source, a compendium of the most salient issues confronting bibliometric analyses, along with a discussion of their likely validity, impact, and limitations.

Data

Analyses in this study were performed on data sets drawn from the Research Evaluation and Policy Project at the Australian National University which has, for many years, worked with citation data obtained from the Institute for Scientific Information (*Bourke* and *Butler*, 1993, 1998). A substantial number of issues faced by this project are broadly generalisable and are clearly similar to those encountered by other research groups in the same area (see especially, *Van Raan*, 1996 and *Martin* and *Irvine*, 1983). This paper will discuss specific methodological issues that have arisen within the context of our Australian experience.

Specifically, analyses in this study are drawn from two ISI derived data sets. The first data set is a collection of all articles appearing in ISI indexed source journals from 1981 to 1995 by authors providing an Australian affiliation (cleaned to the level of department) and the second is an ISI data set of all source and citing articles worldwide in the field of education from 1987 to June of 1998.

The measurement of research quality

The relationship between citations and quality

The initial excitement over the potential of citation analysis was driven to a large extent by the assumption that there is a direct relationship between the number of citations that research receives and the inherent quality of that research (*Wade*, 1975; *Cole* and *Cole*, 1973). Since then, some researchers (e.g. *MacRoberts* and *MacRoberts*, 1996; *Lindsey*, 1989) have pointed out problems with this assumption, but for the most part it still remains generally accepted. One of the most thoughtful discussions of this issue is provided by *Martin* (1996) who carefully distinguishes between research quality, importance, and impact. By thinking of citations as measuring "impact", criticisms that revolve around the notion of "unrecognised" or "innate" quality can generally be avoided.

Unfortunately, regardless of any view those of us who produce citation research may take, and any label we may use, it is still likely that most consumers of citation research

are ready and willing to conceptualise a strong link between citation analysis and the evaluation of research quality. Of course, it is precisely this assumed link to research quality that accounts for widespread interest in bibliometric analysis in the first place.

Certainly a major problem with the notion of quality is that it is difficult to define. Who is to say what it is? It has long been argued that some works of outstanding quality are published that are simply not recognised until years later. Must the quality of an idea be judged in the context of its time, or are good ideas eternal even in the absence of recognition? Not surprisingly, many citation researchers would prefer not to ponder philosophical questions such as these.

Moreover, there is a pragmatic reason to avoid claiming too direct a conceptual link between research quality and citation attainment. Of course, most researchers in the area of bibliometrics realise that citation measures, as useful and as interesting as they may be, are not always (in terms of conception) impeccably precise or completely unambiguous. Unfortunately, consumers of our research, especially government audiences, have a tendency to bestow a greater degree of faith in the preciseness of the measures than is probably justifiable. Using bibliometric measures as an aid for policy decisions is not a problem if one remains aware of exactly what is being measured, and of the related implications. Consumers of citation research, however, are often more focused on broad results rather than on measurement details. Given the complexities of measurement, the concept of *visibility* or *impact* is perhaps a more responsible and less emotionally charged label than research *quality*.

Of course, it has already been shown convincingly that the citations an individual receives strongly correlates with other forms of career recognition, for example, academic position, the attainment of Noble prizes, awards, and membership in scientific academies (see, *Garfield*, 1998; *Zuckerman*, 1977). These relationships also hold in Australia where, for example, less than one percent of Australian university researchers are fellows of the Australian Academy of Sciences, but more than half of the 25 most cited Australian authors have achieved this honour.

The relationship between publications and citations

A simple count of publications clearly provides a very straightforward measure of research output, but might publication counts also be an adequate indicator of the research quality? In fact, there exists a moderately strong correlation between the number of articles a researcher produces and the number of citations attained. Taking Australian educational research as an example, the correlation between the number of papers authors publish and the number of citations they receive is r=0.68. Examining

data for individual authors, however, shows that only seven of the twenty most prolific publishers in education are also in the top twenty in terms of number of citations received. Thus, while it is reasonable to make a general statement such as there is a tendency for individuals who publish a great deal to be highly cited, in any individual case it would be risky to *infer* that because a specific researcher produces a lot of publications they will necessarily also garner a lot of citations.

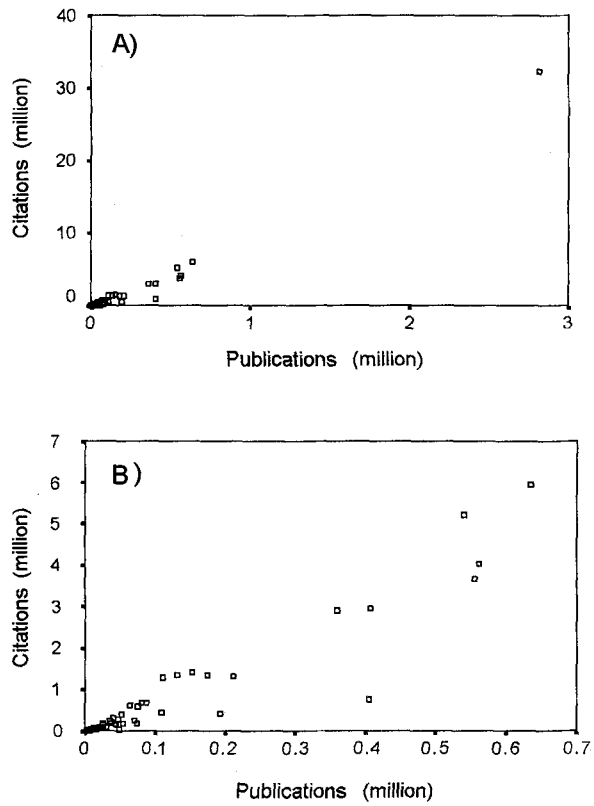

Fig. 1. Citations as a function of publications A) for 90 nations; B) for 89 nations (US ecxluded)

When it comes to more aggregate levels of analysis, however, it does at first appear that publication rates are an excellent indicator of citation attainment. For example, the correlation between number of publications produced and the number of citations

garnered by 90 nations from 1981 through 1994 is r=0.986. Examination of the plot of this strong relationship (see Panel A of Fig. 1), however, does suggest the relationship is rather dominated by a single outlier (the United States). Even when this outlier is removed from the analysis (Panel B) there still remains a strong correlation of r=0.95.

So does this finding imply that at more aggregate levels of analysis, publication and citation counts can be used interchangeably? The answer is no. Considerable caution should still be used when examining individual cases even at high levels of aggregation. Publication and citation data are always highly skewed with a relatively small number of cases accounting for a large proportion of the output. It turns out that even in the case of the very high correlations just reported, the situation of many individual nations would be misrepresented under the assumption that publications are an almost perfect indicator of citations. For example while Sweden ranks thirteenth in publications, it ranks ninth in citations and the former USSR while seventh in publications is thirteenth in citations. India is tenth in publications but nineteenth in citations, and there are many other examples of significant discrepancies. These findings lead to the conclusion that while it is probably acceptable to substitute publication for citation rates for correlational analyses at an aggregate level, it is generally inappropriate to view publication and citation rates as functional equivalents if individual cases are being examined.

Recognition of publications over time

Cole and *Cole* (1973) have examined the issue of article recognition over time. The Coles claim that the number of citations that an article receives in the first few years of publication is an acceptable predictor of the total number of citations it will garner over a much longer time period. On the other hand, *Holton* (1978) claims that very long delays in recognition are quite common. Our Australian data set is of use in considering this issue. Table 1 shows that for articles published between 1981 and 1982, there is a correlation of r=0.87 between citations in 1983-1985 and the total number of citations an article received from 1983-1995. This correlation increases to 0.93 by using a five-year window instead of a three year window. In general, within individual fields the correlations are also high between early and later citations. A notable exception to this is the field of anthropology where the correlation between early and later citation is still below 0.8 even after a five year period. On the other hand, some fields such as psychology, demonstrate a remarkable stability in citation patterns over time. Overall, it does appear that early citations are a good predictor of the number of citations received over a much longer period.

Table 1
Correlations of citations in earlier periods to citations from 1983 to 1995 for articles published in 1981 and 1982 for selected fields

	1983	1983-1985	1983-1987
All Fields	.72	.87	.93
Natural Sciences			
Biological Sciences	.71	.88	.94
Chemistry	.72	.87	.93
Earth Sciences	.68	.86	.93
Engineering	.61	.80	.90
Medicine	.73	.88	.94
Physics	.72	.88	.95
Social Sciences			
Anthropology	.58	.72	.78
Business	.83	.87	.94
Economics	.56	.83	.87
Education	.48	.89	.95
Geography	.56	.81	.91
History	.54	.79	.83
Political Science	.66	.78	.90
Psychology	.87	.96	.98
Sociology	.72	.87	.95

It should be remembered, however, that there are always exceptions to the overall trend. For instance, while we have found a correlation of r=0.94 in the field of medicine between the number of citations a medical paper receives after five years and the number of citations a paper receives after 13 years, this does not mean that the later recognition of any particular paper can necessarily be gauged from citations in the first 5 years. A good example, is an article by *Davies* et al. (1982) that appeared in the *British Medical Journal.* With a total of 226 citations from 1983 to 1995, this paper is one of the 40 most cited papers in the natural sciences during this period with an Australian author. Yet, this paper had only received a single citation by 1985, and only three citations by 1987. What happened was that the article was the first one providing evidence linking a certain antibody to a particular disease, and this work became very highly cited after a new medical diagnostic test was developed. Of course, this sort of situation is an anomaly and early citations are, in general, an excellent predictor of later citations. Nevertheless, it is important to remember that these analytical tools are less than perfect when examining individual cases.

Review and methodological articles

Another criticism of the link between citation analysis and research quality is the claim that highly cited articles tend to be reviews or methodological articles and therefore are not major scientific breakthroughs (*Lindsey*, 1989). It is definitely true that review articles are strongly represented in the set of most highly cited articles. Among our collection of 324,362 papers with Australian authors only 6,685 (or roughly 2%) are review papers. Among the one hundred most cited papers, however, fully 20% of the papers are reviews. And this holds in specific fields like education which is characterised by a one percent incidence of review articles, while 21% of the most-cited articles in education are reviews. There is less evidence of a huge preponderance of highly cited methodological articles in education (only four among the one hundred most cited papers and three of these are also review papers), although it is likely that methodological articles are more characteristic of other fields. It should be remembered, however, that substantial differences exist between fields and a major concern in one field may be irrelevant to another.

The real issue here, however, is why should a methodological or review article that receives many citations not be regarded as being of equal "importance" to other types of highly cited articles? If other researchers find an article to be useful and of sufficient quality to incorporate into their own research, then it would seem that a highly-cited article of this nature makes as important a contribution to future work as any other article that receives an equal number of citations. This is really an issue of semantics. Review articles or statistics texts may not "advance knowledge" directly with new material, but they provide a foundation upon which other scientists can advance knowledge. Articles of this type do, in fact, contribute in important ways to the advancement of knowledge, although this contribution is indirect. This discussion reinforces *Martin*'s (1996) suggestion that the term "impact" might be preferable to more controversial terms such as "quality".

Citation practices

Self-citation

It is generally known that many authors tend to cite their own work. Because of the cumulative nature of individual research, this is not only a natural and acceptable procedure, it is also a useful and informative one. But, how prevalent is this practice and does it invalidate conclusions drawn from data where self-citations have not been excluded?

To consider this question in more detail, a list of the 56 researchers in the field of education with 200 or more citations over a twelve and a half year interval was compiled. The maximum number of citations received by one of these authors was 711, and the average number of citations for this group was 293. The median rate of self-citation was 11.1%. This median rate increases to 13.9% among the ten most highly cited authors, although this is actually a modest rise considering that more highly cited authors also tend to have a larger set of potential publications from which to cite themselves. The correlation between the total number of citations attained by these authors and the total number after self-citations were eliminated was r=0.925.

It is probably the case that self-citation is not a major problem at most levels of analysis. For this phenomenon to influence conclusions in studies of universities or nations, it would have to be argued that the distribution of self-citations is not random, for example, that a particular university or group of universities systematically has authors cite themselves while other universities do not. This seems unlikely.

At the level of individual analysis, however, self-citations could certainly be a significant problem. Only two of the 56 most highly cited authors did not cite themselves and, in the most extreme case, 154 of the 280 citations (55%) a single author received were the result of self-citation. Overall, a consideration of the data suggests that removing self-citations is an important prerequisite when comparing the performance of specific researchers, but that this problem is probably not too serious at more aggregated levels of analysis.

Cronyism

A related concern is that researchers cite their own friends and colleagues and that this practice reduces the value of citation analysis (*Kostoff*, 1998). This practice is sometimes referred to as "cronyism". It is true there are clusters of researchers who tend cite each other a great deal. And it would not be unusual in most disciplines for these people to all know each other at least to some degree. In fact, a group of highly-cited individuals with many mutual citations are quite likely to be "gatekeepers" who form an invisible college (*Crane*, 1972) in a particular field or area. The fact that academic leaders cite each other extensively, however, simply reinforces the fact that attainment of a large number of citations tends to indicate a scholar's position in the field's hierarchy. In other words, when we are considering highly cited individuals, it may well be the case that "cronyism" is little more than a manifestation of the power relations existing within the field. That citation counts reflect this reality is not a methodological shortcoming.

What is productivity?

While it sounds reasonable to identify outstanding and productive researchers in a particular field using bibliometric data, the measurement issues are complicated. Just counting publications is inadequate as some researchers publish a large number of papers, but receive few citations. If citations are counted instead, however, there are other researchers who receive large numbers of citations on a very small number of articles. These may be excellent articles, but it is hard to think of a person who has only produced a couple of articles over a decade or two as highly productive.

Another complication involves authorship position. Should citations be credited to multiple authorship positions or just to first authors? This is likely less of a problem for analyses undertaken at aggregate levels (see *Martin*, 1994), but it makes a difference at the individual level. When we allocated citations equally to multiple authors of articles, we found that the sixth most productive researcher in the field of Australian educational research was a research assistant who was third author on just three papers that turned out to be very highly cited.

For fields in which the most important contributor is generally listed as first author (which is in most, but not all fields) productivity might better be measured by an index that weights both first author publications and citations. In general, an index of this type reflects the attributes that most would agree are highly desirable in a researcher, that is, being both reasonably prolific and able to produce highly visible work. The Productivity Index (*PI*) proposed by *Phelan* (1998) measures this as follows:

$$PI_i = \sqrt{\frac{(p_i \times c_i)}{\sum (p_i \times c_i)}}$$

where p_i equals the total number of first authored publications and c_i equals the total number of citations from first authored publications. If multiplied by 100 for ease of reference, the measure ranges from zero for authors who have no citations to a theoretical limit of 100 in the unlikely case of a single author in a field accounting for all of its publications and citations. This combined index is also useful for measuring institutional or national productivity levels, and at these higher levels of aggregation, the influence of authorship position need not be taken into account. The essential point here is that "productivity" is probably best measured by taking both numbers of publications and numbers of citations into account.

The journal set

The exclusion of journals

While the ISI generally includes among its covered journals, the most highly-cited international journals, there are still substantially more journals in existence than it can include, and obviously many researchers publish in them. This is a particularly important issue for scholars outside of the United States. Australians, for example, publish in many regional journals that are not included in the source index by the ISI. The issue, however, is not that many Australian journals are not indexed by ISI, but whether measures of citations from ISI covered journals are a good indicator of total research activity.

There are two conflicting possibilities. It is possible that certain universities, departments, or individuals produce a substantial amount of work, but focus it in areas not well represented by ISI covered journals, while other institutions and individuals focus on publication in the major international journals that are covered by ISI. An alternative possibility is that institutions and individuals that are productive tend to be productive in multiple areas. This is to say that scholars who publish frequently in ISI journals are likely to be the same ones who publish most in non-ISI journals (and write books, get grants, etc.). Familiarity with the publication process, as well as knowledge of the practices of productive researchers, suggests that the second alternative is definitely plausible.

The Department of Education, Training and Youth Assistance in Australia collects data from universities on all publications in refereed journals whether included in the ISI journal set or not. Using universities as the level of analysis, Fig. 2 shows that there is a strong linear relationship between publication in all journals and publication in ISI covered journals (r=0.967). But will this relationship still hold at less aggregated levels of analysis? In fact, *Bourke* and *Butler* (1996) have shown that there is still a strong relationship between ISI counts of citations and other research activity at the departmental level, even within a single institution, and this relationship is especially strong for the natural sciences. At the same time, it should not be forgotten that there are undoubtedly a few highly specialised sub-fields, especially of regional interest, that the ISI journal set is likely to miss and specialists in these areas are likely not to receive adequate credit of their work through an analysis of ISI data. For Australia, this might be a field like sheep husbandry. For the vast majority of fields, however, ISI data is likely to provide an excellent indication of total research activity.

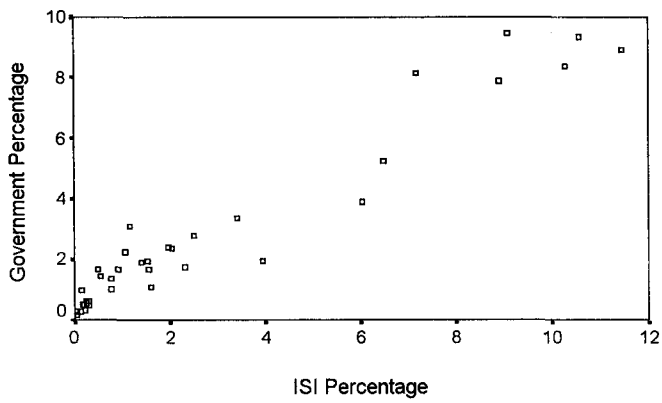

Fig. 2. Percent of publications by government measure plotted on ISI measure of same variable for 35 universities

English language bias

It has long been recognised (e.g., *Carpenter* and *Narin*, 1981) that the ISI journal set is biased toward English language journals. Of course, ISI has been aware of this, and has over the years, greatly strengthened its non-English speaking selection of journals. Also the ISI is continuously identifying journals that receive substantial numbers of citations and adding them to its set of source journals. While there is undoubtedly excellent work appearing, for example, in non-covered Japanese and Russian journals (not to mention classified defence related research), the impact of this research is limited until it finds its way into major international journals, all of which tend to be covered by the ISI journal set (*Cole* and *Phelan*, 1999).

An issue that is similar to English language bias is the idea that scientists in less developed nations are almost systematically being excluded from access to important international journals (*Gibbs*, 1995). This goes beyond the idea that journals in local languages are not catalogued by ISI, to the suggestion that editors of major journals have a tendency to devalue or ignore submissions from these countries. While this is an interesting argument, there is little evidence to suggest this is a widespread phenomenon.

Overall, the problem of non-coverage of journals by ISI does not appear to be a significant one for most analyses, the exception being studies that examine either highly-specialised fields or evaluate regionally important work in non-English speaking countries.

Fluctuations in the journal set

As a matter of routine practice, the ISI journal source index coverage varies quite extensively over time. Many journals start up while others fail and counts of publications over time are definitely affected by how many and which journals are covered (see *Martin*, 1994). As science changes over time, however, it makes good sense that the ISI should change its coverage. The critical question is, how much impact is this ongoing variation of journal coverage likely to have on citation analyses?

In 1981 there were only 6,367 journals covered by our ISI data set, which includes both the social and natural sciences. In that year, Australian authors published in only 37% of these journals. By 1995 journal coverage increased to 8,868 and Australian authors published in 47% of these journals in that year. The increase in the size of the journal set, however, reflects the fluctuation as well as the growth of the journal set. Over the full fourteen year interval covered by our data, 13,569 different journals have been covered by ISI, and Australians have published in 63% of them.

Of the 13,569 journals covered since 1981, only 3,982 (29%) of them comprise the constant journal set by having been covered *every* year. What is of interest is whether findings derived from the constant journal set tend to differ from those derived from the much larger dynamic journal set. It should be noted that the smaller constant journal set does include most of the journals that tend to attract a large number of citations. In fact, of the 1,647,023 citations received by Australian authors from 1981 until 1995, only 18% of them were for source articles not drawn from the constant journal set.

Given that most citations appear in the constant journal set, there is a limit to the extent results can be affected by making use of the dynamic journal set. Still, it is worthwhile to examine the issue more closely. The smaller the unit of aggregation, the more likely there will be a significant impact of journal set selection. Australian educational research at the level of institution provides a useful example as it is not a large area, but it is a rapidly evolving one, as educational methods change frequently and new journals are created to address recent innovations (for example, addressing issues relating to technology and computing in education).

Data were compiled on the number of citations from 1986-1990 and from 1991-1995 received by publications appearing from 1981-1985 and from 1986-1990 respectively. An extraction of these results for the top ten institutional recipients of citations is presented in Table 2.

Table 2
A ranking of Australian universities on number of citations in the field of Educational Research using constant and dynamic journal sets

Panel A: Constant Journal Set

Old Rank		Cites 86-90 (Pubs 81-85)	Cites 91-95 (Pubs 86-90)	Change	New Rank
1	Sydney	110	139	+26.4%	1
2	New England	88	71	-19.3%	3
3	Queensland	67	43	-35.8%	8
4	Monash	58	45	-22.4%	6
5	ANU	55	1	-98.2%	44
6	Macquarie	55	38	-30.9%	9
7	La Trobe	41	32	-22.0%	10
8	Melbourne	34	25	-26.5%	12
9	Curtin	31	51	+64.5%	5
10	James Cook	27	22	-18.5%	14
	All Institutions	775	936		

Panel B: Dynamic Journal Set

Old Rank		Cites 86-90 (Pubs 81-85)	Cites 91-95 (Pubs 86-90)	Change	New Rank
1	Sydney	118	142	+20.3%	2
2	New England	105	72	-31.4%	6
3	Monash	94	102	+8.5%	5
4	Queensland	84	50	-40.5%	9
5	ANU	78	1	-98.7%	46
6	Curtin	59	157	+166.1%	1
7	Macquarie	59	38	-35.6%	10
8	Tasmania	46	24	-47.8%	15
9	La Trobe	41	59	+43.9%	8
10	Melbourne	35	36	+2.9%	11
	All Institutions	1028	1340		

At a single point in time, the correlation for all 54 Australian source institutions for citations received from articles appearing in the constant and dynamic educational journal sets was r=0.968, which suggests that the two data sets are quite comparable. In spite of this strong correlation, however, examination of Table 2 reveals that the exact ranking of individual institutions can be dependent upon the choice of journal set.

Because there is a potential difference in specific findings, it is important to address the *conceptual* question of which journal set is appropriate to use at a particular point in time. Given that the ISI adds journals to its coverage as they start to receive substantial numbers of citations, there is a good argument that the *current* journal set is the appropriate one to examine for analysis at a single point in time. The current journals better reflect contemporary academic interests than any subset can, thus a higher ranking on the dynamic than the constant set indicates an institution that is receiving more recognition in areas that are currently important. While it may appear that ISI is an arbitrary authority of journal importance, it should be remembered that the ISI choice of journal coverage is not guided by whim but rather by which new journals are tending to attract citations. Citations to articles appearing in the current journal set is the best available basis for assessments of performance at a single point in time.

But what about analyses over time? Clearly, some of the change over time might not reflect a change in activity of an institution, but rather just the inclusion of new journals that its researchers have long published in. Regardless, of this possibility, if the argument is accepted that the journal set at a particular point in time best reflects what is going on in a field (and it is difficult to argue that the constant journal set does this better than the dynamic set), then it is still appropriate to examine results from the dynamic journal set *even for analyses over time.* Finally, it should be noted that while choice of journal set can affect the ranking of an individual institution, it will have far less impact on broader correlational analyses, as the relevant measures are very highly correlated with each other.

What is important to remember is that measurement choices definitely can affect outcomes for specific individuals or individual institutions. Thus, any measures provided should be accompanied by a justification as to why the particular measure chosen to examine is appropriate for the analysis. Moreover, any findings regarding individuals or relatively small aggregated entities should be viewed as approximate guides to overall standing, rather than as precise immutable measures.

Technical issues

Identification of authors

While it is true that some names are misspelled, this problem is not so common that it is likely to have a large effect. It is also hard to imagine that the problem of "maiden names" (*MacRoberts* and *MacRoberts*, 1989) is a major one, particularly among highly-cited authors. A more serious problem is that a small but significant number of authors use a different number of initials for different papers, for example, "J. Marcus" on some papers and "J.A. Marcus" on others. Another obvious difficulty is that two or more scholars sometimes have the same name and initials and, of course, the larger the set of authors being considered, the more likely this is to happen. These are all serious problems when considering individuals, but they are not fatal problems. The solution to these problems may be tedious, but it is not impossible. It involves examining every single paper and institutional affiliation of each author being considered. It simply has to be accepted that the identification of individual authors or institutions cannot be accomplished solely by computer analysis. Careful identification techniques can reduce errors to an acceptable level in most cases, although this not always a simple task as some researchers change affiliations more frequently than might be expected. And even with exceedingly careful identification techniques, it is always possible, although presumably rather rare, that two people have the same name in the same department, and do work in the same area. There is no real substitute for consulting with people who are familiar with a field. Still, if the population being examined is not too large, for example, Australian mathematicians, the process of identification is not as difficult as some might think, and is likely to be fairly accurate.

Ambiguity with addresses

Our 1981-1995 Australian data set contains information on more than 300,000 articles written by more than 800,000 authors, and it provides examples of just about every technical problem there is. In our efforts to identify unique institutional addresses, we compiled a list of over 82,000 different addresses that we then amalgamated down to the departmental level. In reviewing these addresses we found over a thousand of them that were either 1) clearly not Australian (several hundred were Austrian, but Germany, Denmark, New Zealand, Scotland, the Seychelles, Poland and numerous other countries were represented), 2) nonsensical (for example, "The University of Queensland in

Melbourne, Victoria") 3) difficult to identify (like "POB 338, Brisbane") or, 4) just sort of amusing (e.g., "The Bra Shop", "The Prostitute's Collective", and the "Australian Camel News").

Even when addresses were clearly related to particular universities, there was still a bewildering array of choices. The University of South Australia, for example, had almost 2000 unique addresses just for it. The Anthropology Department, for instance, is listed as "Anthropol", "Dept Anthropol", "Anthropol Discipline", "Fac Arts, Dept Anthropo", and "Discipline Anthropol". Of course, these data sets are huge, so there is a great temptation to identify groups by computer, but a computer is simply not able to make the necessary choices. In the end, in order to classify 80,000 addresses properly, a person has to go through and review each and every one of them.

Finally, while there is a very reassuring feeling of security when using ISI's excellent compilation software which it provides with its very useful topical citation reports, this software cannot be expected to substitute for detailed local knowledge of what is being examined. For example, the software is, of course, unable to recognise that "Flinders Univ", "Flinders Univ S Australia", "Flinders Unvi S Australia", and presumably "Funders Univ" are all the same institution. The point is that that the ISI staff is confronted with millions of addresses from around the world, many of them not even university addresses, and ISI cannot be expected to sort them all out. This task will always fall to individuals with greater familiarity with local areas and fields. Researchers who obtain data from ISI or other sources and believe they can simply pull out, for example, an accurate list of the twenty most cited institutions without a careful examination of the broader data set, are very likely to produce flawed results.

Nevertheless, while there are definitely serious technical problems with citation data, careful checking, as tedious as it often is, will greatly minimise the probability of inaccurate or misleading results. The problem here rests more with casual research practices than it does with any insurmountable flaws in the data.

Choice of measures

Consumers of citation research often raise many of the issues already discussed in this paper, but our experience is that once placated on these technical issues, they tend to accept, with surprisingly little concern, the details of whatever measure of research performance is chosen for analysis. Most of us who perform citation analyses, of course, recognise that this is a very critical issue. For example, Table 3, provides the ranking of the top five Australian universities in terms educational research compiled using six different measures all delivering somewhat different results. It is not that any of these

six measures are *wrong*, rather it is that each one is reporting a different aspect of publication activity. The trick is to define the problem carefully and choose the most appropriate measure to address it.

Table 3
Six procedures for identifying the five leading Australian Universities in the field of educational research

Panel A: Publications 90-92		Panel B: Citations 90-92	
1. Queensland	7.0%	1. Sydney	11.7%
2. Monash	6.4%	2. Curtin UT	9.2%
3. Deakin	6.0%	3. Monash	7.2%
4. New England	5.9%	4. Deakin	6.3%
5. Sydney	5.5%	5. New England	6.1%
Panel C: Publication/Staff		**Panel D: Citations 90-96 to Pubs 90-92**	
1. W. Australia	1.52	1. Deakin	16.2%
2. Queensland	0.96	2. Curtin UT	7.0%
3. Murdoch	0.74	3. La Trobe	5.4%
4. NSW	0.72	4. Sydney	5.2%
5. Curtin UT	0.65	5. W. Sydney	5.1%
Panel E: Citations 90-96/Pubs 90-92		**Panel F: Pubs 90-92 by Impact**	
1. Deakin	6.6	1. Deakin	71.57
2. W. Sydney	5.5	2. Sydney	68.66
3. Curtin UT	4.2	3. Curtin UT	66.21
4. Wollongong	3.9	4. NSW	65.49
5. La Trobe	3.6	5. Monash	60.29

In most cases, citation measures are likely to be of greater interest than publication measures (but beware, as citation measures, especially at the departmental level, can be greatly influenced by the activities of only one or two lead professors). A simple count of citations (panel B) is interesting, but is a dated measure that might reflect publication activity of long ago rather than more recent activity. Measures of publications (panel C) or citations per staff member are appealing, but it is often very difficult to get accurate counts of staff members who are expected to produce research. The staff counts that panel C is based on (derived from a paper with a vested interest in showing the University of Western Australia in the best light) may be accurate, but are definitely suspect.

Later citations to earlier publications (panel D) are generally a good measure of relatively recent, although not up-to-the-minute activity. Citations per publication (panel E) is also interesting although it is, of course, possible to do very well on this measure with a relatively small output of highly visible publications. The impact measure,

calculated by multiplying each paper by the impact weight (or average number of citations per article) of the journal it appears in, and summing this up over all articles, is useful because it can give an immediate measure of the likely future visibility of recent publications. The problem with the measure is that there tends to be substantial variation in terms of how many citations articles actually receive within the same journal, so results based on impact are not likely to reflect accurately the true level of visibility of a paper over time (which is better shown in panel D). What is clear is that choice of measure can make a substantial difference to the findings emerging from an analysis.

Conclusion

There are numerous issues that should be carefully considered whenever an analysis of bibliometric data is undertaken. It is not clear, however, that there are any more methodological problems confronting bibliometricians than researchers examining many other sorts of data. No data set is completely straightforward. Because critiques of bibliometrics tend to examine multiple levels of analysis, there is an accumulation of identified methodological problems that may or may not be relevant to particular bibliometric investigations at specific levels of analysis. Of course, researchers of bibliometric data need to be aware of these critiques and they need to delineate their assumptions carefully, but by and large, they need not be unduly hindered by most of the caveats discussed in this paper.

The greatest degree of care needs to be taken when examining individual cases. There is always great interest, for example, in ranking tables, whether of individual, institutional, or national performance. There is nothing intrinsically wrong with these, if they are compiled carefully. While rankings accurately report exactly what they purport, there are two problems with them. First, consumers of rankings can hardly restrain themselves from making the conceptual leap from an *indicator of quality* to quality itself. Second, because all we have to offer is indicators, it is likely that while rankings approximate overall reality, they cannot provide pinpoint accuracy for a particular individual case.

Most of the discussion in this paper has focused on measures identifying highly-cited entities. Bibliometrics is best suited for the examination of entities with large numbers of citations. Bibliometric data is, of course, always very skewed with a few cases receiving many citations and the majority of cases receiving very few. There are too many random forces operating to be convinced that there is a real difference between, for example, a researcher who has received only a single citation and another who has received two. But the reality is that a majority of university academic staff never receive

as many as three citations in their whole career. Thus bibliometrics, if it is to claim accuracy, should stay focused on high performers or large aggregates with relatively large numbers of citations. Publication counts are probably a more reasonable measure of performance for the vast lower end of the distribution than are citation counts.

It also deserves mentioning again, that citation and publication counts are two revealing, but still imperfect, indicators of research activity. As *Martin* (1996) has suggested, research evaluation is a field where the use of multiple indicators of research activity is of great value. Bibliometric measures are an outstanding supplement to other measures of research evaluation (for example, peer review), but it would be regrettable if so much confidence was placed in their accuracy that they were routinely substituted for other reasonable methods of performance assessment. Bibliometric analysis is but one tool among many, and so it should remain.

References

BOURKE, P., L. BUTLER (1998), Institutions and the map of science: matching university departments and fields of research, *Research Policy,* 26:711-718.

BOURKE, P., L. BUTLER (1996) Publication types, citation rates, and evaluation, *Scientometrics,* 37:473-494.

BOURKE, P., L. BUTLER (1993), *A crisis for Australian science*, Performance Indicators Project, Monograph Series No. 1, Canberra: The Australian National University.

CARPENTER, M., F. NARIN (1981) The adequacy of the Science Citation Index (SCI) as an indicator of International Scientific Activity, *Journal of the American Society for Information Science,* 32:430-439.

COLE, J., S. COLE (1973) *Social stratification in science*. Chicago:University of Chicago Press.

COLE, S., T. PHELAN (1999) *The scientific productivity of nations, Minerva* 37:1-23.

CRANE, D. (1972) *Invisible Colleges: Diffusion of Knowledge in Scientific Communities.* Chicago: University of Chicago Press.

DAVIES, D. J., J. E. MORAN, J. F. NIALL, G.B. RYAN (1982) Segmental necrotizing glomerulonephritis with anti-neutrophil antibody – possible arbovirus aetiology? *British Medical Journal,* 285:606.

GARFIELD, E. (1998) Random thoughts on citationology. Its theory and practice, *Scientometrics,* 43:69-76.

GIBBS, W. W. (1995) Lost Science in the Third World, *Scientific American.* August: 76-83.

HOLTON, G. (1978) Can Science be measured, in *Toward a Metric of Science.* Y. ELKANA et al. (Eds.) New York: Wiley.

KOSTOFF, R. N. (1998) The use and misuse of citation analysis in research evaluation, *Scientometrics,* 43:27-43.

LINDSEY, D. (1989) Using citation counts as a measure of quality in science: Measuring what's measurable rather than what's valid, *Scientometrics,* 15: 189-203.

MACROBERTS, M. H., B. R. MACROBERTS (1996) Problems of citation analysis, *Scientometrics,* 36:435-444.

MACROBERTS, M. H., B. R. MACROBERTS (1989) Problems of citation analysis: a critical review, *Journal of the American Society for Information Science,* 40:342-349.

MARTIN, B. R. (1996) The use of multiple indicators in the assessment of basic research, *Scientometrics,* 36:343-362.

MARTIN, B. R. (1994) British science in the 1980s: has the relative decline continued? *Scientometrics,* 29:27-56.

MARTIN, B. R., J. IRVINE (1983) Assessing basic research: some partial indicators of scientific progress in radio astronomy, *Research Policy,* 12:61-90.

PHELAN, T. (1998) *Is Australian Educational Research Worthwhile?* Unpublished manuscript. The Australian National University.

VAN RAAN, A. F. J. (1996) Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, *Scientometrics,* 36:397-420.

WADE, N. (1975) Citation analysis: A new tool for science administrators, *Science,* 188:429-432.

ZUCKERMAN, H. A. (1977) *The Scientific Elite.* New York: Free Press.