

General Analyses of Cancer Research Publications in Australian States Using the Science and Social Science Citation Indexes

Concepción S. Wilson

c.wilson@unsw.edu.au

University of New South Wales, School of Information Systems, Technology and Management,
Sydney, 2052, NSW (Australia)

Abstract

This research measures the quantity, quality and extent of international collaboration of cancer research publications in Australian states from 1994-1998 through citation analysis. Journal publications (with at least one Australian author) of the cancer literature from Science Citation Index and Social Sciences Citation Index were analyzed. For the five-year period, New South Wales (NSW) produced the most publications (31%), slightly ahead of Victoria (VIC) with 29%; Queensland (QLD) ranked third (14%) and South Australia (SA) fourth with 11%. However, as measured by mean journal impact factor, the publications from NSW were of overall lower quality than those from VIC, SA, QLD, and from Australia as a whole. When standardized for quality against the national average, and adjusted for state size, the publication output of the four larger states are ranked in order: SA >> VIC >> QLD > NSW. Four measures of international collaboration on publications were investigated as measures of quality; the degree of collaboration with the USA and England broadly support the IF rankings. At a minimum, these results suggest that the relevant Australian State authorities, should initiate more extensive analyses of an apparent deficiency in the overall quality of their states' cancer research, with the view to greater, or more selective, support.

Introduction

This paper is based in part on a study commissioned by the New South Wales Cancer Council (NSWCC) to assess, using scientometric techniques, the quantity and quality of cancer research papers published by researchers in the most populous state of Australia, New South Wales, vis-à-vis those from other states of Australia and especially from Victoria, the second most populous state. Those parts of the resulting report (Wilson & Pittman 2000) both in methods employed and in results obtained, which may have wider interest to the scientometric and informetric community are presented in this paper.

Although numerous studies exist that assess scholarly publications in a variety of science and technology (including medicine) fields, there are only a few that examine or evaluate cancer research publications; these are primarily in European cancer journals assessing the contributions of countries in the European Union (collectively or individually) vis-à-vis that of other countries (Parodi et al 1993; Mela et al 1999; Ugolini et al 2002; Grossi et al 2003; Ugolini & Mela 2003). Other scientometrically related studies on cancer research deal with, for example, the development of ecological oncology (Ketka & Suptelo (1993); the collaborative patterns in the epidemiology of neoplasms (Kundra & Tomov 2001); the negation of citation indexes in assessing quality in lung cancer randomised trials (Berghmans et al 2003); and a meeting abstract on the evolution of breast and ovarian cancer gene research (Bouchard et al 2000).

This study addresses another approach at assessing a sparsely populated (but geographically dispersed) developed country's cancer research performance. It also provides a micro view of Australia's contribution in a specific area of medicine and biology, cancer.

Methods

Census of Publications

A census of the Australian cancer publications in journals was made for each year from 1994 to 1998 inclusive on the Institute of Scientific Information (ISI) databases, *Science Citation Index (SCI)* and *Social Sciences Citation Index (SSCI)*, standard international sources for scientometric analyses. They were searched together through the Dialog Information Services using its duplicate-removal and rank features. All searches were performed in early 2000.

Cancer research in *SCI* and *SSCI* was identified by the presence in subject-related fields (e.g., title, abstract, and descriptors) of keywords covering all aspects of malignant neoplasms. Terms provided in the *Medical Subject Headings (MeSH)* under 'neoplasms', as well as many additional terms were used, thus affording a generous definition of cancer research. This data collection method differs from that using the journal subject category, 'oncology', to select publications on cancer (see, for example, Parodi et al 1993; Mela et al 1999; Ugolini et al 2002; Ugolini & Mela 2003). Publications were treated as Australian if at least one author had an institutional affiliation in Australia. They were identified by the presence of the term 'Australia' in the geographical location field of *SCI* and *SSCI*.

The term 'states' here covers both the six states and the two territories of Australia; reference to them are by their common abbreviations: New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA), Western Australia (WA), Tasmania (TAS), the Australian Capital Territory (ACT), and the Northern Territory (NT). Each publication was allocated to those states in which at least one author had institutional affiliation, with no state receiving more than one allocation per publication. The corporate source field (with institutional and geographic address data) of *SCI* and *SSCI* was used for state identification. Unlike publications from the USA where states are identified through their zip codes that are included in the corporate source field, the equivalent postal codes for Australian states are not included as such. This process needed considerable effort; in the absence of state names or abbreviations, it was necessary to identify full or abbreviated names of major cities and towns in the states; and in their absence, the names of appropriate institutions, e.g., hospitals or research institutes. It should be noted that a single publication would be allocated to more than one state if its different authors listed institutional affiliations in different states. The number of publications so allocated was taken as the research output of each state.

Measurement of Publication Quality

Impact Factors

The term 'quality' denotes (high) conformance to standards. While scientific publications are routinely evaluated with respect to logic of argument, clarity of expression, and so on, the ultimate standard is 'contribution to knowledge', or equivalently, influence or impact on subsequent research through formal citations. To determine this impact a common assumption was adopted; that is, the impact can be gauged by the number of times a publication is cited from a standard set of the literature over some fixed time span. This motivates the use of ISI's annual journal Impact Factor (IF). Formally, this is the number of citations from the ISI databases in a year to all papers in the journal for the two previous years, divided by the number of those papers. That is, IF is a measure of the impact of the 'average paper' in the appropriate issues of the journal on publications in the designated year. The yearly IFs for some 7000 journals are published in ISI's *Journal Citation Reports (JCRs)*. Thus the *JCR* provides a systematic and objective indicator of the relative importance of scholarly journals, and the papers within them, at least (as ISI emphasizes) within a single field of investigation or subject discipline.¹ To proceed we need two further assumptions. First, the impact of any paper in a journal is well-estimated by the impact of this 'average paper', that is, its journal IF – which can be obtained from the appropriate annual issue of *JCR*. Second, a measure of the overall quality of the publications on cancer for each state and each year is given by the mean IF of each set of publications.

International Collaboration

Since the evaluation of the 'quality' of publications is an important component of this study, additional measures to the IF were sought. Apart from evaluation by specialists, this has no recognised substitute,

¹ The IF measure may be better appreciated with examples from the SCI *JCR* for 1997. IFs of nearly 5000 journals range from 0 to ca 41 (citations/publication), but the distribution is very skewed: 96.7% of all journals have IFs of 5 or less, and 63.1% of all journals have IFs of 1 or less; the median IF is only 0.73. In general the higher IFs attach to eminent medical and biochemical journals. Examples, selected from the top 100 journals cited by cancer researchers in NSW and VIC from 1994-1998, are: *Annual Review of Immunology* (37.80), *Nature* (27.37), *Lancet* (16.14), *JAMA* (9.26), *American Journal of Medicine* (4.24), *European Journal of Cancer* (2.41), *Medical Journal of Australia* (1.43), and *Australian and New Zealand Journal of Medicine* (0.56).

and novel but crude measures were devised. It seemed that the degree of international collaboration on publications from a geographical region might be used to gauge their quality (*sensu* influence). The premise is that researchers from one geographical region will seek to collaborate with researchers from another geographical region recognised as more eminent in the same research area, and thereby produce higher quality research. Reciprocal benefits should accrue from collaboration with the better researchers in the less-eminent region. For cancer researchers in the Australian states, therefore, greater collaboration with researchers in, we suggest, the USA and England should indicate a higher proportion of better research publications. This motivates two related indicators of research quality: the percentage of all international collaborations with the USA alone, and the percentage with both aforementioned countries. It also follows that researchers from 'less-eminent' countries than Australia will seek collaborators here, and be drawn more to those regions producing better research publications. This suggests that the total number of international collaborations, and the total number of collaborating countries, both per 100 publications, may also serve as indicators of the quality of the publications of an Australian region. These four variables are merely plausible, not proven, measures, of publication quality; so that their use to supporting or countering IF results is limited. For each Australian publication on cancer, a list of collaborating countries was obtained from its geographic location field. (It should be noted that each country can appear only once in this field per record.) Both the number of collaborations and the identity of each collaborating country were compiled for each state and annual sub-collection, and the required measures directly calculated.

Results

Number of Publications

Australia's share of the World output of journal publications on cancer remained steady at ca.2% over the period investigated (Table 1).² Both the World and Australian annual outputs grew from 1994 to 1997, but (both) declined markedly in 1998. Over the whole period the annual output of Australia grew by 6.1%.

Table 1. The number of publications on cancer for the years 1994-1998, both by Australian researchers and in the World. (Duplicate records, which account for less than 2% of records, have not been removed.)

Year	No. pubs. Australia	No. pubs. World	Australia's share (%)
1994	1660	80425	2.1
1995	1885	86707	2.2
1996	1808	91340	2.0
1997	1959	94294	2.1
1998	1761	84651	2.1
5-yr period	9073	437417	2.1

As explained, inter-state collaboration of authors would result in the publications allocated to a state overlapping to some degree with that of other states. Thus the 'integrity' of the literature allocated to each state must be established prior to making comparisons. Table 2 summarizes data on the allocation of publications: Australia's 8905 unique publications distribute over the states as 10,537 allocated publications; 84.8% of these are from a single state only. Column 5 shows the percentage of those publications allocated to a state which are produced by authors affiliated only with that state. It seems safe to regard the literatures of at least the four top-producing states – NSW (77.4%), VIC (76.0%),

² This is slightly lower than Australia's overall percent share in all fields of nearly 3% (2.83) between 1996 and 2000 (see http://in-cites.com/research/2001/may_14_2001-1.html, Science in Australia, 1996-2000, accessed 8 April 2005.)

QLD (74.5%), and SA (73.2%) – as reasonably independent units for analysis, but, as defined here, other literatures – for example ACT (26.0%) – are more cross-state composites.

Table 2: Details of the allocation of unique publications from Australia over the eight Australian states:

State	Total no. pubs. allocated to state	No. pubs. allocated to one or more other states	No. pubs. from this state only	[No. pubs. from this state only] / [No. pubs. allocated to state] (%)	[No. pubs. from this state only] / [No. pubs. from Australia] (%)
NSW	3252	736	2516	77.4	28.3
VIC	3101	743	2358	76.0	26.5
QLD	1460	373	1087	74.5	12.2
SA	1116	299	817	73.2	9.2
WA	940	391	549	58.4	6.2
ACT	489	362	127	26.0	1.4
TAS	125	50	75	60.0	0.8
NT	54	31	23	42.6	0.3
Total	10537	2985	7552		84.8

9073 publications including duplicates from Australia: SCI/SSCI 1994:1998.

8905 unique publications from Australia: SCI/SSCI 1994:1998. (% of dupls=1.85%)

7552 unique publications from a single state only, or 84.8% of all unique pubs.

1353 unique publications are from more than one state, or 15.2% of all unique pubs.

The number of publications in cancer research as allocated to each Australian state, by year and in toto, for the period investigated, is given in Table 3; Figure 1 displays these data as a percentage of the Australian output for each state. Over the whole period, NSW and VIC were the predominant producers of publications on cancer in Australia, with each state producing a little less than one-third of the national output (Table 3): NSW (30.9%) slightly exceeded VIC (29.4%); the next most productive states are QLD (13.9%) and SA (10.6%).

Table 3: The number of publications on cancer for the years 1994-1998 allocated over the eight Australian states.

Year	NSW	VIC	QLD	SA	WA	ACT	TAS	NT	Total allocation
1994	550	622	254	195	183	85	22	11	1922
1995	633	635	314	261	197	106	26	10	2182
1996	639	634	293	225	182	113	26	8	2120
1997	757	661	320	232	201	94	24	12	2301
1998	673	549	279	203	177	91	27	13	2012
5-yr period	3252	3101	1460	1116	940	489	125	54	10537
(as %)	30.9	29.4	13.9	10.6	8.9	4.6	1.2	0.5	100.0

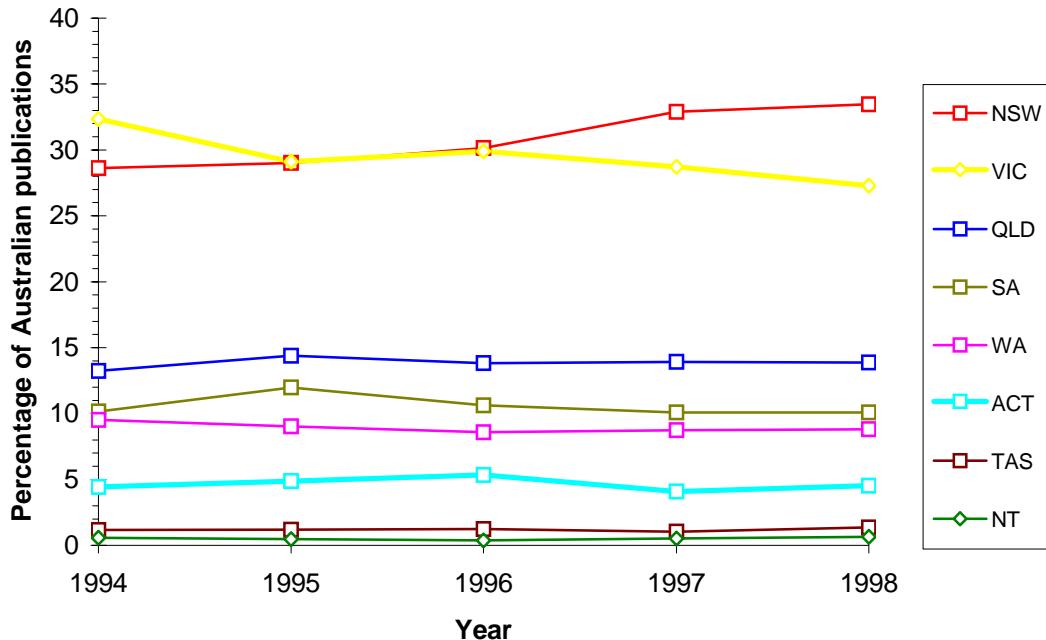


Figure 1: The percentage of Australian publications on cancer for the years 1994-1998 for the eight Australian states.

The annual output of publications from the more productive states roughly follows the national and worldwide trend noted above in Table 1: a moderate growth from 1994 to 1997 with a decline in 1998. In general the percentage of the national annual output remained steady for states other than NSW and VIC (Figure 1). Of most interest: NSW passed VIC as Australia's leading producer of publications on cancer in 1996 and thereafter. At the beginning of the period, the annual output of NSW was below that of VIC (28.6% cf. 32.4% of the national output), but at its end exceeded it by 6.1% (33.4% to 27.3%).

Quality of Publications

Impact Factors

The mean IF for the publications on cancer produced by each Australian state over the period investigated is given in Table 4. These data are displayed in Figure 2: the range of annual mean IFs for each state is shown as a vertical bar, with the five-year mean shown as a horizontal bar. States are placed in decreasing order of their five-year means; thus the mean IFs for the five-year period range from 3.77 for SA to 1.93 for NT. The value of cross-state comparisons of these statistics is weakened by the considerable variation in the annual values, but the comparison of NSW and VIC, each with little variation, is safe. VIC's mean IF not only exceeds that of NSW overall (3.37 vs. 2.69), it exceeds NSW for each year of the study; the minimum annual mean IF for VIC's publications is 3.27 while the maximum value for NSW's publications is 2.88. We also note that VIC's 5-year mean IF lies above the national average of 3.14, while that of NSW lies below it.³

³ A study of oncology research in the European Union (EU) countries as compared to other countries for 1995 – based on oncological journals only – reports a lower mean IF for Australia (2.5); however, this compares favourably with the mean IFs of the EU countries (2.4) and of the World (2.7). In support of the international collaborative measures, note that in this study USA ranks first (3.3) and UK third (2.8), after Canada and the Netherlands (2.9). [See: Mela et al 1999.]

Table 4: Mean Impact Factor for publications on cancer for years 1994-1998 for each Australian state.

Year	NSW	VIC	QLD	SA	WA	ACT	TAS	NT	Australia
1994	2.88	3.27	2.87	3.28	2.37	4.57	2.51	1.53	3.06
1995	2.43	3.48	3.02	4.60	3.06	3.12	1.63	1.22	3.16
1996	2.71	3.40	3.69	3.77	3.33	3.15	1.86	1.71	3.23
1997	2.72	3.35	3.29	3.64	3.47	3.26	2.86	1.39	3.15
1998	2.71	3.34	3.40	3.28	3.26	3.11	2.31	3.45	3.10
5-yr period	2.69	3.37	3.26	3.77	3.10	3.41	2.22	1.93	3.14

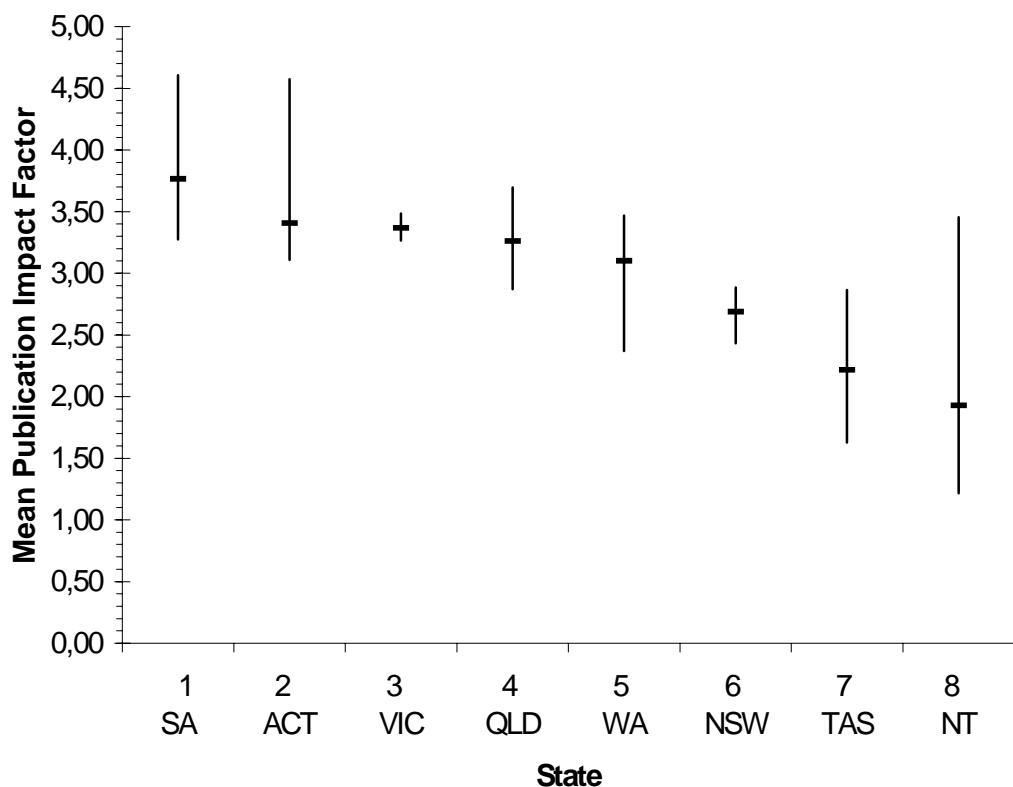


Figure 2: Mean Impact Factor (IF) for publications on cancer for years 1994-1998 for each Australian state.

International Collaboration

Four measures of the degree of international collaboration in the publications on cancer produced by each Australian state over the period investigated are given in Table 5. Two of these measures, the mean number of international collaborations, and the mean number of (non-Australian) collaborating countries, both per 100 publications, proved to be dependent on the total number of publications produced by a state; this is most apparent for the second measure (Column 6). Without further correction for the effects of size, we are limited to comparing states with similar publication outputs, and fortunately NSW and VIC qualify. For both measures over the five-year period, VIC does slightly better than NSW: 52.5 vs. 47.4 collaborations on 100 publications, and 2.6 vs 2.4 collaborating countries on 100 publications. The other two measures, the percentage of all collaborations with the USA, and with the USA and England, seem size-independent, and are strengthened by the USA

invariably ranking first and England (with one exception) ranking second throughout. The values for each measure have been assigned a ranking for decreasing order over the states. On the first measure, VIC is ranked first with 28.8% of its international collaborations involving the USA, while NSW is ranked sixth with 22.5%. On the second measure, VIC is again ranked first, with 40.3% of its international collaborations involving the USA and England, while NSW is ranked fourth with 36.8%. In summary, on all four measures, publications on cancer from VIC exceed that of similar publications from NSW for the five-year period 1994-1998.

Table 5. Four international collaboration measures for the cancer literature for each Australian state for 1994-1998:

State	No pubs	Collaborations		Collaborating countries		Percentage of collaboration with top country (USA)		Percentage of collaboration with top 2 countries (USA & England*)	
		no.	/ 100 pubs	no.	/ 100 pubs	%	rank	%	rank
NSW	3252	1540	47.4	79	2.4	22.5	6	36.8	4
VIC	3101	1629	52.5	80	2.6	28.8	1	40.3	1
QLD	1460	886	60.7	59	4.0	25.1	4	36.1	5
SA	1116	624	55.9	51	4.6	25.6	3	38.0	3
WA	940	528	56.2	53	5.6	27.8	2	39.2	2
ACT	489	300	61.3	45	9.2	24.0	5	34.0	6
TAS	125	52	41.6	19	15.2	19.2	8	30.7	7
NT	54	40	74.1	17	31.5	20.0	7	30.0	8

Number of international collaborations per 100 publications,

Number of non-Australian collaborating countries per 100 publications,

Percentage of collaborations with top collaborating country (USA),

Percentage of collaborations with top two collaborating countries (USA and England*),

*For TAS, England is the third ranked collaborating country.

Discussion

The results show that in the period of study NSW and VIC each produced about one-third of Australia's publications in cancer in Australia, with NSW's output slightly larger; QLD produced about one-seventh, and SA and WA about one-ninth. However, as measured by IF, and generally supported by additional measures, the NSW output overall is of lower quality than that of either SA, VIC, QLD or WA.

Further useful contrasts may be drawn by transformations of the basic data. First, the quality and quantity measures may be combined and standardized by multiplying the number of publications per state by the mean IF of publications per state, and dividing by the national mean IF, for the five-year period (Table 6). In terms of such 'equal-quality publications', the output of the four larger states may now be ranked, in two subgroups, as follows: VIC > NSW >> QLD > SA. Second, some compensation for differences in size of the states as to numbers of researchers, research budgets, etc., can be achieved by dividing the previous results by states' population (Table 6).⁴ In terms of equal-quality publications per 10,000 people, the output of the four larger states shows an altered ranking, in three subgroups, as follows: SA (9.2) >> VIC (7.5) >> QLD (4.9) > NSW (4.6). We note that two size-independent collaborative measures proposed (percentage collaboration with USA, and with USA and England) support the ranking: VIC > SA > QLD > NSW.

Any contention regarding this study likely centres on the labelling of some states' research as of lower overall 'quality', so it is appropriate to briefly review our use of IFs. Our working assumption is

⁴ Population data are taken from the 1997 revised concise edition of *The Times Atlas of the World*, London: Times Books.

that the quality, in the sense of influence, of a paper in a journal relative to other papers in the same area of research, is adequately estimated by the mean number of citations per paper to that journal from the ISI databases in the two years following its publication. We first draw the reader's attention to criticisms of the journal IF as defined, e.g., that the citing databases are biased to English-language sources; that the chosen period for papers to receive citations may be distortingly short for many research fields; and that the mean is an unsuitable statistic of the citedness of the typical journal paper, given the strongly skewed nature of the distribution of citations to papers in a journal (Moed et al 1999; Seglen 1992).

Table 6: The publication output on cancer research for Australian states for 1994-1998, standardised with respect to quality, and corrected for state size.

State	No. pubs	Mean IF	Publication output standardised by quality		Standardised publication output corrected for state size		
			(No. pubs x Mean IF) / Aust mean	Rank	Population (1000s)	Stand. pubs per 10,000 people	Rank
NSW	3252	2.69	2786	2	6009	4.6	6
VIC	3101	3.37	3328	1	4462	7.5	3
QLD	1460	3.26	1516	3	3113	4.9	5
SA	1116	3.77	1340	4	1462	9.2	2
WA	940	3.10	928	5	1678	5.5	4
ACT	489	3.41	531	6	299	17.8	1
TAS	125	2.22	88	7	472	1.9	8
NT	54	1.93	33	8	168	2.0	7
Australia	10537	3.14			18054	5.8	

Of immediate interest is whether there is a misleading bias with respect to states in our use of IFs. For instance, is the mean IF an unsuitable statistic for comparing state IF distributions? (Might VIC have a few very highly cited papers each year vis-à-vis NSW, perhaps from one research team, for example?) The annual distributions for the four top-producing states show a similar form, so that though skewed, their central tendency is adequately represented by the mean for present purposes; i.e., a lower mean IF systematically reflects a higher proportion of publications in low IF journals. Might inter-state bias arise in estimating each paper's IF with its journal IF? This could only be answered by an extensive study of the citations received by individual papers, but it does seem unlikely with averaging over large data sets and for high-producing states with many separate sources of research publications. A related problem may be perceived in the plausible case of appreciable differences between states in the overall subject focus of their cancer research, with major differences in the citedness of the associated journals. (Might research in VIC vis-à-vis NSW be more highly focussed on heavily-researched biochemical problems, for example?). The argument might then be advanced that it is unfair to label research in small sub-fields as lower quality simply because small sub-fields generate fewer citations; but we reiterate that, in the sense of impact on the total body of cancer research used here, this would indeed be the case. Finally, it may be argued that differences in mean IFs – e.g., between NSW and VIC, roughly between 2.7 and 3.4 citations per average paper – are too small to warrant the significance we have attached; in defence, we note that this NSW-VIC difference persists through each of the five years of the study.

At a minimum we suggest these results should give the relevant state authorities pause, and prompt further analyses of what seems to be a problem in the quality of some research. A stronger conclusion is that additional, or more selective, investment in cancer research is required in some states. As this study represents aggregate data of the Australian states' publication in cancer research, it would be necessary to explore a lower level of aggregate data; for example, the institutional sector such as universities and hospitals, to determine if the differences observed in the indicators are due to

more basic or clinical interests of the researchers or specialities. A similar study for the next five-year period (1999-2003) could be conducted both at the macro (states) and micro (institutions and/or specialities) levels of analyses to see if there have been some improvement in the quality of cancer research undertaken by researchers and clinicians in Australia.

References

Berghmans T.; Meert A.P.; Mascaux C.; Paesmans M.; Lafitte J.J. & Sculier J.P. (2003). Citation indexes do not reflect methodological quality in lung cancer randomised trials. *Annals of Oncology*, 14, 715-721.

Bouchard L.; Dalpe R. & Ducharme D. (2000). The evolution of breast and ovarian cancer gene research: a bibliometric study. *American Journal of Human Genetics*, 67, 511.

Dialog Thomsom Information Services. (2005). [<http://www.dialog.com>, accessed 8 April 2005.]

Journal Citation Reports (1994:1998). A bibliometric analysis of science (social sciences) journals in the ISI database. Philadelphia, PA: Institute for Scientific Information.

Grossi F.; Belvedere O. & Rosso R. (2003). Geography of clinical cancer research publications from 1995 to 1999. *European Journal of Cancer*, 39, 106-111.

Ketko E.G. & Suptelo L.P. (1993). An analysis of modern trends in the development of ecological oncology (scientometrics of information flows). *Ekperimentalnaya Onkologiya*, 15, 74-79.

Kundra R. & Tomov D. (2001). Collaboration patterns in Indian and Bulgarian epidemiology of neoplasms in Medline for 1966-1999. *Scientometrics*, 52, 519-523.

Medical Subject Headings: Main headings, subheadings and cross references used in the Index Medicus and the National Library of Medicine Catalog. Bethesda, MD: National Library of Medicine. [<http://www.nlm.nih.gov/mesh/>, accessed 8 April 2005.]

Mela G.S.; Cimmino M.A. & Ugolini D. (1999). Impact assessment of oncology research in the European Union. *European Journal of Cancer*, 35, 1182-1186.

Moed H.K.; Van Leeuwen T.N. & Reedijk J. (1999). Towards appropriate indicators of journal impact. *Scientometrics*, 46, 575-589.

Parodi S.; Parodi A.; Lombarndo C. & Santi L. (1993). Cancer research in the European community and other non-EU countries. *Tumori*, 79, 9-15.

Rodrigues P.S.; Fonseca L. & Chaimovich H. (2000). Mapping cancer, cardiovascular and malaria research in Brazil. *Brazilian Journal of Medical and Biological Research*, 33, 853-867.

Science Citation Index. (1994:1998). Philadelphia, PA: Institute for Scientific Information.

Seglen P.O. (1992). The Skewness of Science. *Journal of the American Society of Information Science*, 43, 628-638.

Social Sciences Citation Index (1994:1998). Philadelphia, PA: Institute for Scientific Information.

Ugolini D.; Casilli C.; Mela G.S. (2002). Assessing oncological productivity: is one method sufficient? *European Journal of Cancer*, 38, 1121-1125.

Ugolini D. & Mela G.S. (2003). Oncological research overview in the European Union: A 5-year survey. *European Journal of Cancer*, 39, 1888-1894.

Wilson C.S. (1999). Informetrics. *Annual Review of Information Science and Technology*, 34, 107-247.

Wilson C.S. & Pittman S. (2000). *Assessments of outputs: Quantity and quality of cancer research publications in New South Wales from 1994 to 1998*. Sydney: The University of New South Wales, School of Information Systems, Technology and Management.