Refereed papers

Health informatics research in Australia: retrospective analysis using PubMed

Kumara Mendis MBBS MSc (Medical Informatics) MD (Family Medicine)
Senior Lecturer, School of Rural Health, University of Sydney, Australia

ABSTRACT

Objective To obtain an overview of Australian health informatics (HI) research through a bibliometric analysis using PubMed.

Method Australian HI publications from 1970 to 2005 were downloaded in Medline format using PubMed queries. These were written to a Microsoft Access database using a software application, PubMed Grabber/Analyzer, developed by us. Search Query Language and online PubMed queries were used for further analysis.

Results Publications increased from three (1970) to 335 (2005), with the rate increasing since 2002. Medical Journal of Australia (177), Australian Family Physician (66) and Australasian Physical and Engineering Sciences in Medicine (64) are the top three journals publishing HI articles. Coiera EW (21), Metcalfe P (19), Ebert MA (17), Kron T (16) and Westbrook JI (16) were the five most frequent authors. Of the 2350 total publications categorised according to PubMed publication types, there were 231 reviews, 137 clinical or randomised controlled trials, 64 letters, 44 editorials and 12 meta-analyses. From 1990 to 2005, Australian HI publications in PubMed increased 10-fold (34–335), compared with the total world HI publications, which increased from 2777 to 15 353. It also compared favourably with other medical discipline publications: general practice (69–346), cardiology (440–928) and public health (1779–5724).

Discussion Australian HI publications in PubMed increased consistently throughout the period 1970–2005, which is on a par with world trends in HI publications. In Australia, most HI publications are in general medical journals. Lack of consistency in author names and failure to include the country name and even the state are the main obstacles to PubMed bibliometric analysis.

Keywords: bibliometrics, medical informatics, PubMed

Introduction

Health informatics (HI) is defined broadly as ‘the study of how we organise ourselves to create and run healthcare organisations to deliver the best possible healthcare’.1 In the era of evidence-based medicine,2 HI is essential to bridge the research–practice gap that keeps widening with exploding medical knowledge; human brain power cannot keep pace with this explosion.

Worldwide, developed countries have taken initiatives to invest in and promote HI to improve health care. The United Kingdom National Programme for IT in the National Health Service (NHS) may be the biggest health informatics project undertaken to connect more than 30 000 general practitioners and 300 hospitals in a single country.3 Australian initiatives, such as the HealthConnect4 and MediConnect5 projects, plan to connect electronic medical records across the country using a secure system in order to improve quality and safety.

The establishment of the National Electronic Health Transition Authority (NEHTA)6 and the identification of SNOMED-CT7 as the preferred national terminology in Australia have been important recent milestones.

Research is the driving force of science. In 1999, the Australian government decided to double health and medical research (HMR) expenditure every four years.8 This brings HI into the forefront even more, as biotechnology and genetic research9 will be top priorities. Furthermore, HI will be the key component
that will enable the benefits to be delivered directly to the Australian population with the least possible delay.

Bibliometric analyses are used to keep track of research. Analysing research publications using PubMed in 2005, we found that compared with clinical disciplines like general practice and cardiology, the emerging field of medical informatics has achieved impressive research publications. The only bibliographic study on health informatics in Australia we managed to retrieve was a book review. Our objective in this study was to obtain an overview of Australian HI research through a bibliometric analysis using PubMed.

**Method**

PubMed is the web interface of Medline, the largest biomedical bibliographic text database in the public domain. It contains more than 16 million citations and abstracts from 4800 journals, published in 70 countries, from the fields of medicine, nursing, dentistry, veterinary medicine, the healthcare system and preclinical sciences. **Entrez** is the integrated, text-based search and retrieval system used for the major databases, including PubMed and other molecular biology and genomic and taxonomy databases. PubMed was used to retrieve the number of HI publications from Australia from 1970 to 2005.

A PubMed publication consists of many so-called 'tags' that are abbreviated names for the different fields. The tags are structured in a way that is similar to a structured abstract of a journal article, but they are more comprehensive. The main fields that we used in this analysis are author affiliation (AD), publication type (PT) and medical subject heading (MeSH).

The AD tag is the field that includes the institutional affiliation and address of the first author. We used the information in this field to obtain the publications from Australia. We searched for the word 'Australia' and also for all state and territory names in the AD tag. The AD tag was also used to count the publications originating from universities. The MeSH tags are the main keywords of PubMed and the number currently stands at 22,997. Specificity can be increased when MeSH words are used to query PubMed.

In formulating queries to retrieve publications from Australia, our aim was to increase sensitivity to minimise missing publications that originated from Australia. However, this may increase false positives. For example, the American state of Washington (WA), or Hong Kong (NT), will be retrieved if the abbreviations for 'Western Australia' and 'Northern Territory' are used. We minimised the false positives by including Boolean NOT statements to exclude such countries.

When retrieving HI publications, we erred towards specificity as there are MeSH words specific for most of the HI-related publications. We used eight MeSH words: Informatics, Medical Informatics Applications, Medical Informatics, Medical Informatics Computing, Nursing Informatics, Dental Informatics, Public Health Informatics, Computational Biology. The only TEXT word used was 'Health Informatics'. PubMed is queried by entering a word, phrase or a combination of words joined by Boolean terms. The Entrez system automatically translates these words or phrases into the query language. Alternatively, it is possible to enter the exact query into the PubMed interface; this is what we opted to do. The main PubMed queries that we used are listed below. The LIMIT was to retrieve a date range or publication type, etc.

**Query 1** – to obtain all Australian publications:


data


**Query 2** – a query that captured health informatics publications:


data


**Query 3** – by combining queries 1 and 2 (using AND), all HI publications by Australian authors were obtained. Similarly, combining each country (UK, New Zealand) with health informatics gave the respective countries' publications.

Subdomains of health informatics used by MeSH – dental, nursing, public health and bioinformatics (mapped to computational biology) – were searched to determine the subgroups of articles. Primary care informatics, an emerging subdomain that is particularly relevant to Australia, was searched as a 'text word'. Query 3 was applied to each subdomain.

The results of the PubMed queries can be downloaded in various formats. We used the Medline 'TEXT' format. Using a software application (PubMed Grabber/Analyzer [PGA]) developed at the Computer Centre, Faculty of Medicine, University of Kelaniya, Sri Lanka, the Medline text files were written to a relational database management program in Microsoft
Access. The development and application of the PGA software has been reported previously.\textsuperscript{17} The frequency lists of authors and journals were obtained by querying the Access database with Search Query Language (SQL). These were then cross-checked with online PubMed queries.

We retrieved the data from PubMed for this paper during April 2006.

Results

There were 2350 Australian HI publications from 1970 to 2005. Publications increased from 37 in 1990 to 335 in 2005; see Figure 1. The rate doubled from 151 to 318 within a span of four years (2000–2004). However, the increase from 2004 to 2005 was only 17. This may be because all the publications for 2005 may not have been indexed in Medline by April 2006 when the data were retrieved. The corresponding publications in the UK increased from 150 to 463 and in New Zealand from 19 to 69. The Australian HI publication growth seems to be higher than the ‘ALL Countries’ increases (533 to 14,493) during the study period. For both Australia and the UK, there has been a greater increase in publications commencing around 2002, with a narrowing in the gap between the two countries in 2005. On a per capita basis, probably Australia outperforms the UK.

Of the top five journals with HI publications in Australia, the leader is the Medical Journal of Australia (\textit{MJA}) (see Table 1). Only three journals have a ranking with the Journal Citation Index in 2005, with Physics in Medicine and Biology at sixth position the highest with a rating of 2.683.

EW Coeira led the author list with 21 publications (see Table 2), followed by P Metcalfe (19), MA Ebert (17), JI Westbrook (16) and T Kron (16). There were 15 authors with 10 or more publications, but the vast majority (80\%) had only a single publication.

While HI publications increased from 37 to 335 from 1990 to 2005, general practice publications increased from 92 to 425, cardiology from 440 to 928 and public health from 1779 to 5724 (see Table 3).

Categorised according to the PubMed publication type, we found 231 reviews, 137 clinical or randomised clinical trials, 44 editorials and 12 meta-analyses.

University, department or faculty-generated (as stated in the author affiliation) publications numbered 1243 (54\% of the total).

The respective numbers of articles in the subdomains of informatics are as follows: bioinformatics – 189, public health informatics – 17, nursing informatics – 15 and dental informatics – 1. There were 38 primary care informatics-related articles which were retrieved using the [ALL] tag. Of the total, 28 were retrieved as a result of the authors’ affiliations to departments of primary care informatics.

Discussion

Australian HI publications in PubMed increased consistently throughout the period 1990–2005, but markedly since 2002. HI research in Australia is on a par with world publication trends, but may even surpass
the UK on a per capita basis. As an emerging discipline, HI compares well with other established medical disciplines.

The domain of bibliometric analyses is widening. It is being used to study trends in fields like epidemiology,18 and the National Health and Medical Research Council (NHMRC)19 uses publication analysis as one of the key indicators both in the short and medium term to track research grant outcomes. It has been used to track narrow domains such as nursing informatics or even social sciences.

In Australia, HI articles are published mainly by non-informatics journals like the Medical Journal of Australia, Australian Family Physician, Australian Health Review and Australian and New Zealand Journal of Public Health. However, in a study about ‘evaluation studies of information technology in health care’, Ammenwerth & de Keizer20 noted a ‘strong shift from medical journals to medical informatics journals’. The top 10 contained only two specialised journals of informatics. There were only 11 publications in the Journal of the American Medical Informatics Association.

<table>
<thead>
<tr>
<th>Table 1 The 10 top journals publishing Australian health informatics indexed in PubMed from 1967–2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

* ISI Web of Science, Journal Citation Index

<table>
<thead>
<tr>
<th>Table 2 Authors with 10 or more health informatics publications in PubMed from 1970–2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4=</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8=</td>
</tr>
<tr>
<td>13</td>
</tr>
</tbody>
</table>
(JAMIA), which is one of the highest ranking informatics journals (JCI – 2.612). It is worthy of note that Coiera or Westbrook contributed to eight of the JAMIA articles.

One limitation with our analysis relates to incomplete or incorrect author affiliations, sometimes omitting the country or state. We agree that hand-searching would be the most accurate way of searching the literature at this point in time; however, online searching is much more cost and time-effective. Even searches for Cochrane reviews are a combination of online and hand-searching.

Generating the author table was problematic. This is due to the lack of consistency in publishing author names to journals. Some journals require complete first names while others require various combinations of the surname and initials. Use of one consistent surname and the same initial(s) by authors would be of immense help in generating an accurate author table. We searched using state names in addition to using the country name (Australia) because of incomplete entry of the country name in the AD tag. However, if for example the AD tag was searched for Victoria, authors from Victoria, Canada would also be retrieved. Using only the MeSH tag for countries (geographical location) will decrease the sensitivity.

Our study analysed HI publications generated from Australian institutions, and not from all Australian researchers anywhere in the world. For example, if an Australian researcher worked for an institution in the United Kingdom and published papers, these could not be counted with our method.

Not all Australian HI articles might be indexed in PubMed. A study that compared medical informatics articles in two different databases, Medline and LILACS, concluded that ‘Medline properly represents the impact of medical informatics in non-Latin American international journals, but lacks a considerable amount of articles from this region, while LILACS, although in comparison it is smaller in size, has more articles regarding the subject.’ This could be due to the problem of Spanish being the main language in Latin America. In Australia this argument cannot be valid as Australian publications have a high visibility in PubMed and English is the main language of publication. However, the technical publications related to health and medical informatics that are published in

### Table 3: Australian PubMed publications in health informatics compared with general practice, cardiology and public health from 1990–2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Health informatics</th>
<th>General practice</th>
<th>Cardiology</th>
<th>Public health</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>37</td>
<td>92</td>
<td>440</td>
<td>1779</td>
</tr>
<tr>
<td>1991</td>
<td>48</td>
<td>97</td>
<td>455</td>
<td>2078</td>
</tr>
<tr>
<td>1992</td>
<td>53</td>
<td>137</td>
<td>502</td>
<td>2342</td>
</tr>
<tr>
<td>1993</td>
<td>74</td>
<td>137</td>
<td>554</td>
<td>2467</td>
</tr>
<tr>
<td>1994</td>
<td>64</td>
<td>151</td>
<td>572</td>
<td>2467</td>
</tr>
<tr>
<td>1995</td>
<td>69</td>
<td>167</td>
<td>575</td>
<td>2723</td>
</tr>
<tr>
<td>1996</td>
<td>83</td>
<td>195</td>
<td>573</td>
<td>2877</td>
</tr>
<tr>
<td>1997</td>
<td>116</td>
<td>179</td>
<td>577</td>
<td>3286</td>
</tr>
<tr>
<td>1998</td>
<td>129</td>
<td>230</td>
<td>675</td>
<td>3492</td>
</tr>
<tr>
<td>1999</td>
<td>150</td>
<td>252</td>
<td>708</td>
<td>3655</td>
</tr>
<tr>
<td>2000</td>
<td>151</td>
<td>247</td>
<td>691</td>
<td>4082</td>
</tr>
<tr>
<td>2001</td>
<td>173</td>
<td>293</td>
<td>846</td>
<td>4568</td>
</tr>
<tr>
<td>2002</td>
<td>186</td>
<td>290</td>
<td>846</td>
<td>4982</td>
</tr>
<tr>
<td>2003</td>
<td>257</td>
<td>359</td>
<td>944</td>
<td>5517</td>
</tr>
<tr>
<td>2004</td>
<td>318</td>
<td>417</td>
<td>949</td>
<td>6018</td>
</tr>
<tr>
<td>2005</td>
<td>335</td>
<td>425</td>
<td>928</td>
<td>5724</td>
</tr>
</tbody>
</table>
journals such as Decision Support Systems, and IEEE journals, might not get into PubMed.

Currently, ‘Health Informatics’ is not an MeSH term in PubMed. It does not even automatically map to the closest relation – ‘Medical Informatics’ – in the way that ‘General Practice’ maps to ‘Family Practice’. Primary care informatics has been proposed as a subdomain of medical informatics and we found 38 articles related to it. However, they were indexed using the MeSH headings ‘Medical Informatics’ or ‘Primary Health Care’. Given the proliferation of subdomains and also emerging parallel domains such as bioinformatics, clarification regarding the relationships with the core discipline of medical informatics would be helpful in the development and the correct use of bibliographic knowledge bases.

The growth of Australian HI publications in PubMed compares well with the total worldwide HI publications, and also with publications from other Australian medical specialties. HI publications are on a par with world publication trends, but maybe even surpass the UK on a per capita basis.

ACKNOWLEDGEMENTS

I would like to thank Professor Rick McLean for assistance with editing and Professor Michael Kidd for advice regarding historical facts of health informatics in Australia.

REFERENCES

3 National Programme for IT in the NHS. www.connectingforhealth.nhs.uk
5 MediConnect. www.mediconnect.gov.au
7 SNOMED-CT. www.snomed.org/snomedct/index.html
22 Cochrane Collaboration. TSC [Trials Search Co-ordinators] and hand-searching resources. www.cochrane.org/resources/hsearch.htm

CONFLICTS OF INTEREST

None.
ADDRESS FOR CORRESPONDENCE:

Kumara Mendis
Senior Lecturer
School of Rural Health
University of Sydney
PO Box 1043
11 Moran Drive
Dubbo, NSW 2830
Australia
Tel: +61 2 6885 7977
Fax: +61 2 6885 7979
Email: kmendis@med.usyd.edu.au

Accepted February 2007