Three decades of the *Ceylon Medical Journal*— analysis using MEDLINE (PubMed)

Kumara Mendis[1], Indragith Solagarchchi[2] and **Chaminda Weerabaddana[2]**

(Index words: citation analysis)

Abstract

Introduction The *Ceylon Medical Journal (CMJ)* is the only Sri Lankan medical journal that is indexed in MEDLINE (PubMed). Publications in the *CMJ* from 1965 March to 2001 December were analysed using the PubMed interface of MEDLINE.

Method Using PubMed we downloaded the *CMJ* bibliography in MEDLINE format. Important tagged fields were written to a Microsoft-Access database using a PubMed Grabber/Analyser program. The analysis was done using Access-SQL and PubMed queries.

Results There were 1472 citations by 1373 authors. 944 authors had only one and 176 had two articles. The top 10 authors contributed 8% and the top 25, 15.4% of the articles. Publications types were: 68.5% 'journal articles', 10.4% letters, 4.2% historical articles and 4.1% reviews. Controlled clinical trials, randomised controlled trials and clinical trials together totalled only 39 (2.1%). Articles were classified using 2 to 44 medical subject headings (MeSH; average 11.4, mode 9) from the MeSH vocabulary. *CMJ* articles classified using broad MeSH categories were (top five): infections 370(15 %), pathological conditions signs and symptoms 266(10.8%), haemic, lymphatic and immunologic 199 (7.7%), endocrine, nutritional and metabolic 189 (7.7%), neoplasms 179(7.3%).

Discussion For 938 (68%) authors, publishing in the *CMJ* was a one time affair. The top 50 authors contributed nearly a quarter (23.2%). A product of this research is an off-line *CMJ* searching system from 1965 to 2001 with menu driven search facilities which will be a useful tool for researchers.

[1]Head, [2]Demonstrator, *Computer Centre, Faculty of Medicine, University of Kelaniya, Sri Lanka.*
Correspondence: KM, *e-mail: <kmendis@sltnet.lk> (Competing interests: none declared). Received 5 June 2004 and revised version accepted 10 November 2004.*

Introduction

The *Ceylon Medical Journal (CMJ)* was established in 1887 as the publication of the Ceylon Branch of the British Medical Association. It ceased publication in 1893 and was revived in 1904 under another name. The journal continued uninterrupted thereafter and in 1952 resumed its old name of the *Ceylon Medical Journal* [1]. It is the oldest surviving medical journal in Australasia, and the only Sri Lankan medical journal that is indexed in MEDLINE, the largest medical electronic bibliographic database in the public domain.

MEDLINE is produced by the National Library of Medicine (NLM) in USA. NLM is manned by more than 40 full-time indexers, which includes content as well as information science experts [2]. They add more than 500 000 biomedical article citations to MEDLINE yearly, which contain more than 12 million citations from 4800 journal articles in life sciences with a concentration on biomedicine.

Index Medicus was the National Library of Medicine's first bibliographic publication after its establishment in 1879 by the physician-librarian collaboration [3]. *Index Medicus* is the indexing system for finding the articles published in biomedical journals. Incidentally, the *CMJ* was first published a few years after this, in 1887. The *Index Medicus* was computerized in 1960 and was known as MEDLARS. The facilities for on-line searching MEDLARS using a telephone connection commenced in 1971. 'MEDlars on-LINE' was known as MEDLINE. Later in the 1990s, Hilary Clinton, then America's first lady, announced worldwide free access to Medline through the internet called PubMed (www.ncbi.nlm.nih.gov/entrez/).

When a journal article is indexed in MEDLINE, keywords are added by the NLM indexing staff. They assign 10–12 most appropriate MeSH terms (commonly known as keywords or index words) to a journal citation. These keywords come from medical subject headings (MeSH). MeSH is used to describe what an article or book is "about." That is, as index terms they provide an indication of the major topics under consideration [4]. Applying the MeSH vocabulary ensures that articles are uniformly indexed by subject, whatever the author's words. In 2002, there were about 19000 MeSH words in use.

Bibliographic publications about Sri Lankan medicine cover the period from 1811 to 1988 [5–7]. We could not find any in-depth citation analysis of the *CMJ* using paper or electronic bibliographic sources. Doing citation analysis manually is expensive and time consuming. Therefore we decided to make use of MEDLINE to analyse the *CMJ*. Currently, MEDLINE citations are available from 1965 onwards, and is the main reason why our analysis was limited, from 1965 to 2001.

Method

MEDLINE is available as an on-line resource on Internet, free of charge at PubMed - www.ncbi.nlm.nih.gov/entrez/. We downloaded from PubMed the *CMJ* citations for the years 1965 to 2001. To download the citations we used the search words *"Ceylon Medical Journal"* in the input box of PubMed and a limit criteria from 1965 to 2002/12/31. The detailed search string was ("Ceylon Med J"[Journal] AND ("1965"[PDat]: "2001/12"[PDat])). The data we downloaded was in MEDLINE citation format, which is a text file that can be read using any word processor such as Microsoft Word. A MEDLINE citation consists of many fields of information (e.g. title, author, source etc.) that is similar to a structured abstract of a journal article plus other structured information (e.g. MeSH words, country, author affiliation, etc) that is included by the NLM staff at the time of indexing.

Using a software application (PubMed Grabber/Analyzer-PGA) [8] developed at the Computer Centre, Faculty of Medicine, Ragama, the important fields in the MEDLINE text file was written to a relational database management program in Microsoft Access 2000. The citation fields that we captured were: unique identifier (UI), publication ID (PMID), title (TI), authors (AU), affiliation (AD), source (SO), date of publication (DP), publication type (PT), country of origin (CY), Medical Subject Headings (MH), Abstract (AB). The analysis of this database was done using our program—PGA.

The results that we obtained using PGA were crosschecked with direct on-line PubMed queries wherever it was possible to do so, for example (a) author frequencies of the top five authors individually, (b) the publication types for *CMJ,* and (c) the disease categories. The disease categories were tabulated mainly using on-line PubMed queries. We tried to match the disease categories for hospital morbidity and mortality given in the Annual Health Bulletin (AHB), 2000 [9]. Our objective was to categorise the articles (using the MeSH words) into the groups that resembled the ICD-10 main groups given in the AHB 2000.

A paper on the methodology and application software of PubMed Grabber/Analyser is published elsewhere [8]. Although we completed our research in October 2002, all web sources were accessed again in March 2004.

Results

There were 1472 citations by 1373 authors for the period 1965 March to 2001 December. Each citation (or an article in a journal) had one or more authors. The number of articles per year (Figure 1) have been increasing from 1965 to 2001 (Range 15 – 74, Mean 39.7, SD 17.1). The highest increase (76 to 112) in the number of articles published was seen during the years 1991–1992 and 1993–1994, leading to the highest number (74) in 1998. Only 37 (2.7%) authors had 10 or more and 124 (9%), four or more articles. The majority 944 (69%) had only a single article (Table 1).

Articles were classified using 2–44 keywords (Mean 11.4, Mode 9) from the MeSH vocabulary. MeSH uses keywords classify a citation into publication types, subject matter or content, etc. (Table 2)

Figure 1. **Number of articles published in *CMJ* from1965–2001**

Table 1. **Number of articles published with percentage of authors**

Number of articles	Number of authors	% authors
1 – 4	1248	90.9
5 – 9	87	6.4
10 – 15	25	1.8
16 – 20	7	0.5
21 – 32	6	0.4

Table 2. **Frequency of publication types**

Publication type	Number	%
Journal article	1249	68.5
Letter	189	10.4
Historical article	76	4.2
Review	75	4.1
Review, tutorial	60	3.3
Biography	43	2.4
Comment	40	2.2
Editorial	27	1.5
Clinical trial	24	1.3
Randomised controlled trial	10	0.5
News	7	0.4
Bibliography	5	0.3
Review of reported cases	5	0.3
Controlled clinical trial	5	0.3
Review, multicase	3	0.2
Multicenter study	2	0.1
Congresses	1	0.1
Meta-analysis	1	0.1
Evaluation studies	1	0.1
Total	**1823**	

Note: there was one or more publication type for a single citation

Publications types (Table 3) were mainly journal articles 1249 (68.5.%), letters 189 (10.4%), historical articles 76 (4.2%) and reviews 75 (4.1%). Controlled clinical trials, Randomised controlled trials and clinical trials together totalled only 39 (2.1%).

Articles classified under broad disease categories using the MeSH words are given in Table 3. The top ten categories were: infections 370 (15 %), pathological conditions, signs and symptoms 266 (10.8%), haemic, lymphatic and immunologic diseases 199 (7.7%), endocrine, nutritional and metabolic diseases 189 (7.7%), neoplasms 179 (7.3%), environmental disorders 176 (7.1%), nervous system 152 (6.2%), cardiovascular diseases 151(6.1%), digestive diseases 150 (6.1%), and congenital, hereditary and neonatal diseases and abnormalities 133 (5.4%). All other categories comprise 20.6%.

Table 3. **CMJ articles categorised according to main MeSH disease categories using PubMed search filters**

Rank	Disease categories in MeSH	Number	%
1	Bacterial infections, mycoses, parasitic diseases, viral diseases	370	15.0
2	Pathological conditions, signs and symptoms	266	10.8
3	Haemic and lymphatic diseases, immunologic disease	191	7.7
4	Endocrine diseases, nutritional and metabolic diseases	189	7.7
5	Neoplasms	179	7.3
6	Disorders of environmental origin	176	7.1
7	Nervous system diseases	152	6.2
8	Cardiovascular diseases	151	6.1
9	Digestive system diseases	150	6.1
10	Congenital, hereditary and neonatal diseases	133	5.4
11	Female genital diseases and pregnancy complications	129	5.2
12	Skin and connective tissue diseases	127	5.2
13	Respiratory tract diseases	100	4.1
14	Urologic and male genital diseases	74	3.0
15	Eye diseases	18	0.7
16	Otorhinolaryngologic diseases	17	0.7
17	Others	43	1.7
Total		**2465**	**100**

Discussion

There was a general trend of increasing number of articles in the *CMJ*, particularly since 1988. For the majority (944) of the 1373 authors, publishing in the *CMJ* was a one-time endeavour.

Out of the 1472 citations, only 2.6% of the articles were of the MeSH category 'clinical trials.' [4] Analysing 40 years of randomised trials in the *New Zealand Medical Journal,* it was found that a large proportion of published randomised trials are not correctly identified on MEDLINE. Therefore the exact percentage of trials in the *CMJ* if searched by hand could be more than 2.6%. Some argue that hand searching of entire journals remains the most accurate method for identifying the eligible reports, but nevertheless is the most time-consuming [5]. This can be rectified to a great extent if the authors submit the most important keywords that would reflect the contents of an article. If the scope

of the field we are interested in is broad (e.g. internal medicine or an entire journal) and the number of articles huge, the manual searching option becomes less feasible. One other explanation may be that most Sri Lankan authors publish clinical trials in foreign journals. We hope to address this question in an article that will focus on all Sri Lankan medical publications indexed in MEDLINE.

We wanted to ascertain how the content of articles published in the *CMJ* reflected the Sri Lankan morbidity and mortality patterns. Categories such as infections, disease of environmental origin (Injuries, poisoning) the top ranks in both the *CMJ* (Table 3) and the AHB [9]. Diseases of the respiratory system, conditions originating in the perinatal period, and pregnancy, childbirth and puerperium although high up on the AHB morbidity and mortality rankings, were not so well represented in *CMJ*. Our attempt was not easy because of the different classifications used. The Sri Lankan hospital morbidity and mortality statistics [9] are routinely coded using the WHO specification, which is currently the International Classification of Diseases (ICD – 10th revision). The MeSH keywords that are used to index MEDLINE journal articles are a part of the Universal Medical Language Systems (UMLS) meta thesaurus, which is closely related to the Systematized Nomenclature of Medicine–SNOMED coding system. Even the top hierarchies in the two systems, ICD-10 and MeSH, terms are not compatible even at the broadest categories. Therefore we could not properly reconcile and compare the percentages of the MeSH disease categories of the *CMJ* with that of the ICD-10 categories of the AHB.

PubMed / MEDLINE has been used in many different ways. Some used it as in the present study, to analyse articles in a single journal published over several years [6], others studied the change in MeSH words to examine the varying patterns in American medicine. All these studies were done using PubMed Internet interface and on-line queries. One reason may be that on-line searching is a standard feature available in developed country academic settings. Nevertheless, a properly working Internet connection which will allow a user to search PubMed is a luxury even in most academic institutions in Sri Lanka. In a country like Sri Lanka where telephone call rates are extremely high, doing analysis using on-line PubMed will cost a lot of time and money.

The software system we developed—(PubMed Grabber / Analyzer) [8]—the offline software program for searching the *CMJ* will be useful to scientists in selecting relevant researches. It will simplify browsing through the journal without incurring an additional cost of telephone call charges and Internet surfing. By downloading the new citations at regular intervals, off-line searching facilities of *CMJ* can be kept up to date. This *CMJ* analysis software will be available at the Sri Lanka Medical Library for use of doctors in the future.

Similar analyses can be done using commercial bibliographic management software like EndNote (http://www.endnote.com) or Referrence Manager (http://

www.refman.com). These have to be purchased from the web sites paying US $ 50–100 for a single user license. Even if one has an authorized version, a simple analysis such as the 'frequency of all authors in *CMJ*' may not be possible with the normal menu driven commands of Reference Manager.

Acknowledgements

References

1. Uragoda C G. A history of medicine in Sri Lanka –from the earliest times to 1948. Colombo. Sri Lanka Medical Association. 1987. Middleway Limited.Colombo 2.

2. National Library of Medicine. Facts Sheet. Bibliographic Services Division http://www.nlm.nih.gov/pubs/factsheets/mesh.html (Accessed on 30–03–2004).

3. Schulman J, Using Medical Subject Headings (MeSH) to examine patterns in American medicine. [course paper, STS 5206]. Falls Church: Virginia Polytechnic Institute and State University, Northern Virginia Center; May, 2000 http://www.nlm.nih.gov/mesh/patterns.html (Accessed on 30–03–2004).

4. Nelson S J, Johnston D, Humphreys B L Relationships in Medical Subject Headings. In: Bean Carol A, Green, Rebecca, editors. *Relationships in the Organization of Knowledge*. New York: Kluwer Academic Publishers, 2001: p: 171–184. http://www.nlm.nih.gov/mesh/meshrels.html (Accessed on 30–03–2004).

5. Peiris K, Uragoda C G. *Bibliography of Medical Publications relating to Sri Lanka 1811–1976*. Colombo. National Science Council of Sri Lanka.1980: 453.

6. Uragoda C G. *A Bibliography of Medical Publications relating to Sri Lanka 1981–1988*. Colombo, Author Publication. 1983: 175.

7. Peiris K, Uragoda C G. *A Bibliography of Medical Publications relating to Sri Lanka 1977–1980*. Colombo: Ceylon College of Physicians,1983: 164.

8. Mendis K, Solagarchchi I, Weerabaddana C. PubMed Grabber / Analyzer – a tool for off-line analysis of PubMed. 2003. Proceedings of the 10th Asia Pacific Association of Medical Informatics Conference. Deagu, Korea.

9. Ministry of Health. Annual Health Bulletin 2000. Colombo. Page 85

10. Neal B, Rodgers A, Mackie MJ, MacMahon, S. Forty years of randomised trials in the New Zealand Medical Journal. *New Zealand Medical Journal* 1996; **109**:372–3

11. Jadad AR, MqQay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online Journal of Current Clinical Trials* 1993 Feb 27; Doc No 33.

12. Simo Minana J, Gaztambide Ganuza M, Latour Perez J. [Atencion Primaria Journal in MEDLINE: an analysis of the first 7 years of indexing (1989—1995)]. *Atencion Primaria* 1999; **23** (Suppl 1): 5–13 [Article in Spanish].