Evolution of medical informatics in bibliographic databases

Paula Otero, Federico Pedernera, Sergio Montenegro, Damian Borbolla, Sebastián Garcia Marti, Marcela Martinez, Daniel Luna, Fernán Gonzalez Bernaldo de Quirós

Department of Medical Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Medical informatics became a medical specialty during the last years and this is evidenced by a great amount of journal articles regarding the subject published worldwide. We compared the presentation of Medical Informatics in two different bibliographic databases: MEDLINE and LILACS (Latin American and Caribbean Literature on the Health Sciences). Previous studies described how Medical Informatics was represented in MEDLINE, but we wanted to compare it to a regional database as LILACS. We search both databases completely (MEDLINE 1966 -2002 and LILACS 1982-2002) using the keyword "Medical Informatics" as MeSH term in MEDLINE and as DeCS term in LILACS, and we added "medical informatics" as text word and analyzed the references obtained as results. We found that MEDLINE properly represents the impact of Medical Informatics in non-Latin-American international journals, but lacks of a considerable amount of articles from this region, while LILACS, although in comparison it is smaller in size, has more articles regarding the subject. So we think that LILACS properly represents the specialty in Latin America and the Caribbean Region.

Keywords:

Medical Informatics; Periodicals; Subject headings; Information Storage and Retrieval; Databases, Bibliographic; Latin America; Caribbean Region.

Introduction

Medical Informatics (MI) is an interdisciplinary speciality supported by different related areas as research; medical education; clinical practice; information sciences and informatics.

The main feature of MI is the development of a new model regarding the use of bio-information. This is characterized by the production of structures whose aims are the correct presentation of knowledge; the perfection of the information use according to processes and actors of the former processes; the creation of data retrieval and layout mechanisms; and finally the integration of different sources of information.(1)

The field of MI has grown during the last years (2), a fact evidenced by the diversity of international specialized publications in English language regarding the specialty. However, this presence has not been observed in Latin American publications yet.

The National Library of Medicine (NLM) defines Medical Informatics as "The field of information concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine". Its presentation in MEDLINE has changed according to the evolution of the controlled vocabulary called "Medical Subject Headings" (MeSH), used by the National Library of Medicine (NLM). (3)

The term "Medical Informatics" was added to the MeSH thesaurus in 1987. From 1982 to 1986 articles were indexed as "Information Systems", and from 1966 to 1981 as "Information Retrieval Systems".

Previous analysis have assessed different aspects of medical informatics presentation by international literature, using as the main reference MEDLINE bibliographic database.(4-8) Therefore, the main reason of the present study is the analysis of MI representation in the biomedical Latin American literature, by means of LILACS (Latin American and Caribbean Literature on the Health Sciences) and MEDLINE databases.

LILACS

LILACS database (9) is maintained by BIREME (Biblioteca Regional de Medicina – Regional Library of Medicine, Sao Paulo, Brazil). It includes health sciences literature published in the Latin American region from 1980. It gathers the joint effort of more than 400 Cooperative Centers from 27 countries of Latin America and the Caribbean region. LILACS includes articles from about 630 renowned health journals; with almost 300,000 entries (September 2002) classified as original articles; thesis; chapters from books; reports from congresses and conferences ; scientific and technical reports and government publications.

DeCS (Spanish acronym for Descriptors in Health Sciences) (10) is the controlled vocabulary structured used for the indexation of the citations presented in three languages DeCS (11) was developed from MeSH "Medical Subject Headings"(3) of the U.S. National Library of Medicine . The objective of DeCS is the uniform use of terminology in three languages, being a unique and consistent source of independent information retrieval in English, Portuguese and Spanish. Due to the diversity of documents stored in LILACS, we decided to evaluate each group separately. The groups were Monographs (including unusual documents, projects and conference proceedings); periodicals (LILACS term for journal) which

are similar to the bibliographic citations indexed by MEDLINE database; and thesis.(12)

The impact factor (IF) (13)of the main publications was also measured in order to establish their features.

In Table 1 we summarize LILACS database features and in Table 2 there is a comparative chart between MEDLINE and LILACS

Table 1: LILACS database features (updated data 09/2002) (14)

Entries	**297,815** (76% periodicals; 18% monographs; 4% thesis; 2% not conventional)
Language	55% Spanish; 37% Portuguese; 7% English
Countries	**27** (Brazil 51%; Chile 8%; Argentina 7%; Mexico 6%)
Citations per year	**12400** (min. 1980: 690 - max. 1997: 19932)
Titles per country	**618** (Brazil 41%; Argentina 17%; Chile 10%; Mexico 6%; Venezuela 5%)

Table 2: Comparative table between LILACS (14) and MEDLINE (15)

	LILACS	MEDLINE
Date ranging	1982-+	1966-+
Entries	300,000	+12,000,000
Type of publication	Journal articles, Monographs	Journal articles, Reviews, Thesis
Journals indexed	618	4600
Countries	27	70
Controlled Vocabulary	DeCS	MeSH
Topics	Health Sciences	Biomedical research, clinical sciences
Main Language	Spanish 55%	English 89%

Materials and Methods

In July 2002 a search of keywords or MeSH terms was performed in the bibliographic reference database MEDLINE by means of the interface PubMed. The key words or MeSH terms were: Medical Informatics; Medical Informatics Applications; Medical Informatics Computing; Decision Making; Computer-Assisted; Information Storage and Retrieval; Information Systems; Decision Support Systems; Clinical; Hospital Information Systems; Integrated Advances Information Management Systems; Management Information Systems; Medical Records Systems and Computerized. In order to avoid the loss of related articles, the phrase "medical informatics" was added to the title and abstract fields of the citations. The date range analyzed was from 1966 to June 2002. In October 2002 LILACS bibliographic database was searched for the available data on the Web by means of BVS. The search used the same terms because the DeCS terms used

by LILACS are identical to the MeSH terms. The whole database was analyzed (from 1982 to October 2002).

Due to the different structures of the former databases; their features were analyzed separately.

In the search performed in MEDLINE, we analyzed from the whole citations: the key words; the different types of publications included; the source countries and languages and the scientific journals where the publications were included.

In the search performed in LILACS, we analyzed from the whole citations: the key words or descriptors used (major or main and minor or secondary); the different types of publications included; the source countries and the scientific journals where the articles were published. We classify them into periodicals, monographs and thesis.

The impact factor (13, 16) of the different publications of the group "periodicals" (considering year 2000 as a reference) was measured in both groups.

Results

MEDLINE

The search performed in PubMed (17) in July 2002 found 33196 bibliographic citations that suited the search parameters. PubMed interface incorporates many bibliographic databases; apart from those included in the Index Medicus. The analyzed citations are distributed among the following subsets of information: Index Medicus (59.3%); Health Management – HealthSTAR (28.7%); Nursing (6.4%); Health Technology (3.2%) and 2.4% distributed among dentistry, history of medicine, aerospace medicine, biotechnology and bioethics.

MeSH terms used

The NLM indexation method gives a MeSH term to every citation in order to classify them. The average was 8 MeSH terms for each citation resulting from the search, ranging between 1 and 32.

The most frequent MeSH terms were: Electronic Medical Record (17.4%); Hospital Information Systems (16.3%); Medical Informatics (5.5%); Internet (5.1%); Decision Making; Computer-assisted and Information Storage and Retrieval, both with 3.8%.

A great diversity is evidenced probably due to the vocabulary indexation organization.

Publications

Citations analyzed regarding MI came from 3262 scientific publications, from which 2762 (84.7%) had less than 10 articles published. The group of publications that included more than 100 citations published gathered 44 journals (1.4%).

It is worth of mention that 71.6% of the citations are summarized in 20 publications, most of which are specialized journals or summaries from MI congresses. Only 6 of the former journals have IF, and the BMJ is the only general medical publication of this group. (Table 3).

The presence of MI was analyzed in those publications with an IF of more than 3. This is summarized in 11 publications with a total of 1997 articles (6%). Publications in order of importance according to the IF are: New England Journal of Medicine; Nature; Science; JAMA, Lancet; Annals of Internal Medicine, Nucleic Acids Research, BMJ, Radiology, American Journal of Public Health and Journal of the American Medical Informatics Association.

Table 3 : 15 first publications on MI in MEDLINE (number, Impact factor, percentage)

Name of the publication	n	IF	%
Healthc Inform	901	nd [1]	6.6
Stud Health Technol Inform	781	nd	5.8
Methods Inf Med	697	0,929	5.1
Medinfo	621	nd	4.6
Proc Annu Symp Comput Appl	609	nd	4.5
Comput Healthc	592	nd	4.4
Health Manag Technol	528	nd	3.9
Nucleic Acids Res	506	5,396	3.7
Health Data Manag	468	nd	3.5
J AHIMA	461	nd	3.4
Mod Healthc	460	nd	3.4
Proc AMIA Symp	435	nd	3.2
Bull Med Libr Assoc	361	nd	2.7
J Am Med Inform Assoc	354	3,089	2.6
J Med Syst	349	nd	2.6
MD Comput	343	nd	2.5
Hospitals	335	nd	2.5
Int J Med Inf	317	0,699	2.3
BMJ	305	5,331	2.2
Med Inform (Lond)	283	1,182	2.1

[1] Not defined

Type of publication

Citations were included among 31 different types of publications. 90,8% of the citations were original articles, 5% were revisions, 0.9% were controlled randomized studies plus multi-centric studies and meta analysis.

Language

Citations were distributed among 27 different languages. The prevailing language was English (89.3%); followed by German (2.7%), French (1.8%); Russian (1.6%); Japanese (0.9%) and in the fifth place Spanish (0.63%).

Countries

Bibliographic references came from 64 countries. Most of the citations related to MI came from the U.S.A. (61%); followed by the United Kingdom (14.6%); Holland (5%); Germany (4.9%); and Canada (2.9%).

The Latin American region represents only 0.3% (113 citations) of the total. If we add publications from Spain the percentage increases to 0.8%. Latin American citations are headed by Brazil (50.4%); Mexico (18.6%); Chile (13.3); Argentina (9.7%); and the rest is divided among Venezuela; Puerto Rico and Cuba.

Evolution per year

Citations related to MI have increased during the last years, the date range 1966-1981 represents 10.9%; 1982-1986 (10%); 1987-2002 (79,1%). Citations of the last ten years represent 63.3% of the whole citations; which shows the advance of the specialty during the last decade.

LILACS

Taking into account the results obtained from LILACS database in October 2002. The search strategy retrieved 1825 bibliographical citations. As we explained before, LILACS database indexes different kind of documents. The citations found had the following percentages: monographs and non conventional documents (48%); periodicals (47%); and thesis (5%).

Citations came from 20 different countries from the region and the full text may be read in electronic format in 9.6% of them (CD-ROM or Internet).

Subject Descriptors (DeCS) used

LILACS assigns several DeCS terms to every reference in order to classify them properly. These DeCS or keywords are categorized in major or main and minor or secondary. We found 1155 main descriptors and 886 secondary descriptors in the retrieved citations.

The main descriptors most frequently used were: information systems (13.7%); medical informatics (6%); computer programs (3.2%); computer systems (1.6%); epidemiological surveillance (1.3%); Internet, computing in medical informatics, medical informatics use, automatic data processing, and computers networks shared 1.1%.

The most frequent secondary descriptors were: information systems (6.6%); brazil (3.6%); computer programs (3.6%); medical informatics (2.1%), Latin America (1.5%); America (1.3%); handbook (1.2%) and medical informatics applications (1.1%).

A great diversity is evidenced probably due to the vocabulary indexation organization.

Periodicals ("journals")

The citations regarding MI analyzed in the "periodicals" group are included in 298 scientific publications; 284 (95.3%) of which had less than 10 articles published. The publication that indexed more citations was the report from the "Pan American Sanitary Bureau" (Oficina Sanitaria Panamericana) with 22 references. The analysis of the main publications is detailed in Table 4.

From the former group, the only publication that had impact factor (IF) was the "Revista Médica Chilena" with 0.290 and 16 citations regarding MI. The publications which had less

than 10 citations and which had impact factor (year 2000 database) were: Genetics and molecular biology (7 citations and 0.470 IF); "Revista de investigación clínica" (1 citation and 0.202 IF); "Acta bioquímica clínica latinoamericana" (2 citations and 0.043 IF)

A great diversification of the publications exists and we could not find in the database a specialized publication or a set of scientific journals presenting MI in a complete way.

Table 4: Main publications containing articles related to MI in LILACS

Summarised title	Number
Bol. Oficina Sanit. Panam	22
ACIMED	21
Rev. saúde pública	21
Rev. Bras. Inform. Saúde	20
Divulg.saúde debate	17
Rev. Méd.Chile	16
Inf. Epidemiol. SUS	15
Rev. Chil. Pediatr	13
Evidencia aten. Primaria	13
Rev.Hosp Matern. Infant.Ramón Sardá	12
Arq. Bras cardiol	12
Mundo saude	12
Brasilia med	11
Acta med Hosp. Clin Quir Herman Ameijeiras	11

Countries

The bibliographic references from 1825 citations were from 20 different countries. It is very important to relate the former percentages with the number of journals indexed by country, because every country and its cooperative centre is responsible for the classification and indexation of the national journals. From October 1992 every country is responsible for the selection of their journal titles according to the following criteria: Scientific content; Peer Review; Editing Commission; Punctuality of the publication; Frequency; Duration; Standardization and Graphic layout.(14)

The Latin American country that publishes more citations related to MI (according to the number of indexed publication) is Uruguay. The average of the former country is 7 articles per indexed journal (June 2002 database). Table 5 details the relationship between the percentage of MI articles and the number of indexed journals per country.

The high percentage of the U.S.A. is outstanding, this could be related to the number of publications indexed and written by the Pan American Health Organization (PAHO). Even though it only indexes 3 publications, a great amount of its database material came directly from the PAHO.

Table 5: Relation between MI citations / number of indexed journals

Country	Citations	Articles / Journals per node
Brazil	797	3.2
U.S.A.	283	141.5
Argentina	139	1.3
Chile	118	2.0
Cuba	88	2.6
Peru	71	3.9
Mexico	70	1.8
Uruguay	56	7.0
Costa Rica	46	3.3
Bolivia	35	2.3

Evolution per year

Citations related to MI published in scientific periodicals (series periódicas) have increased as the database evolves . In 1982 there were only three citations. In 2000 the peak was 81 citations. The average is 42 citations / year and there is a growing tendency.

Discussion

Last years´ development of MI is evidenced by international literature. A tendency to report the use of Electronic Medical Records and Hospital Information Systems is observed. The cause of the former situation is that PubMed interface incorporates MEDLINE citations as well as HealthSTAR database (Health Services; Technology; Administration; and Research). The latter database focuses on clinical and non clinical aspects of health care, analyzing the patients therapies results, the effectiveness of the procedures, programs, services, administration and planning. MI is not yet a topic frequently analyzed in general publications. Only 2.4% of the citations were published in one of the "5 greats": BMJ; New England Journal of Medicine; Lancet; JAMA; Annals of Internal Medicine. Most of the articles are presented in congresses rather than in renowned journals, and a great number of them will not be published in an international publication. The cause of the former situation is difficult to understand but this is also evidenced in other medical specialties.

We can agree with Morris (5), who states that there are no more than 30 publications related to MI.

LILACS database has evolved and at present it is consider one of the main bibliographic databases when doing a complete bibliographic search. It is important to use it when doing a systematic review. Many search strategies have been proposed, which shows LILACS´ development as a bibliographic database during the last years. (18, 19)

We can not find data related to the whole available Latin American biomedical publications. However, we feel LILACS represents the publication from the region.

Medical Informatics advance has been evidenced worldwide by international literature, however Latin American representation regarding the subject is not clear. We could not find many specialized publications. From the journals that presented more citations related to MI, only the "Revista Brasileira de Informatica em Saude" was a specialized publication. The former journal was indexed only for the year 1988, and the others were general medical publications or medical specialized journals.

The diversity of documents included in LILACS (monographs, thesis, conference proceedings, original articles, etc) enlarges the number of publications. But if we take as a bibliographic parameter an original article (equivalent to LILACS "periodicals"), only 47% of the citations were periodical publications.

There was also a difference between the keywords found on both databases, we think that this could be attributed to the fact that there almost no medical informatics specialized journals on LILACS, and another view could be that the assignation of keywords in MEDLINE is done by specialized staff of the NLM, while in LILACS each participating node is responsible to assign the keywords, this could lead to a different criteria when assigning keywords.

The Impact Factor (IF) (13, 16) is an assessment tool created by the "Institute for Scientific Information © (ISI ©)" by one of its functions devoted to the citations analysis , known as "Journal Citation Reports" © (JCR ©). It is useful when analyzing periodical publications and users´ specific needs. The publications analyzed that had a higher IF were generally not clinical journals. Only 4 publications from the 300 journals assessed in LILACS had IF published based on the year 2000 data. The Latin American publications IF is not high, due perhaps to the little diffusion of our regional publications worldwide.

Almost 10% of the citation offer access to the full text in an electronic format, which acts as a divulgation tool of the scientific activity of the region.

We found that only four publications (of the main one containing articles related to MI) overlapped when comparing MEDLINE to LILACS, so we can say that very few Latin-Americans journals are indexed in MEDLINE

Conclusion

The presentation of MI in scientific journals in our region has not been properly used yet. However, LILACS bibliographic database accurately represents the information published in the region.

References

[1] Lorenzi NM. The cornerstones of medical informatics. J Am Med Inform Assoc 2000;7(2):204-5.
[2] Power M. An approach to policy analysis and development of medical informatics. Methods Inf Med 1999;38(4-5):260-4.
[3] Coletti MH, Bleich HL. Medical subject headings used to search the biomedical literature. J Am Med Inform Assoc 2001;8(4):317-23.
[4] Moorman PW, van der Lei JL, van Bemmel JH. An analysis of 5 years of medical informatics publications. Medinfo 1998;9 Pt 1:140-4.
[5] Morris TA, McCain KW. The structure of medical informatics journal literature. J Am Med Inform Assoc 1998;5(5):448-66.
[6] Sittig DF. Identifying a core set of medical informatics serials: an analysis using the MEDLINE database. Bull Med Libr Assoc 1996;84(2):200-4.
[7] Sittig DF, Kaalaas-Sittig J. A citation analysis of medical informatics journals. Medinfo 1995;8 Pt 2:1452-6.
[8] Vishwanatham R. Citation analysis in journal rankings: medical informatics in the library and information science literature. Bull Med Libr Assoc 1998;86(4):518-22.
[9] Literatura Latinoamericana y del Caribe en Ciencias de la Salud. In: BIREME; 2002.
[10] BIREME. DeCS - Health Sciences Descriptors. In: BIREME.
[11] Jimenez Miranda J. Acceso a MEDLINE y LILACS mediante el MeSH y el DeCS. ACIMED 1998;6(3):153-62.
[12] Pediatría. SA. Manual de LILACS. In: Sociedad Argentina de Pediatría; 2002.
[13] Garfield E. The impact factor. In: Institute for Scientific Information; 1994.
[14] BIREME. Metodología LILACS. In: BIREME; 2002.
[15] NLM. Medline Fact Sheet. In: NLM - National Library of Medicine.
[16] Garfield E. Using the impact factor. In: Institute for Scientific Information; 1994.
[17] Medicine NLo. Entrez database - PubMed. In: NLM - National Library of Medicine; 2002.
[18] Castro AA, Clark OA, Atallah AN. Optimal search strategy for clinical trials in the Latin American and Caribbean Health Science Literature database (LILACS database): update. Sao Paulo Med J 1999;117(3):138-9.
[19] Clark OA, Castro AA. Searching the Literatura Latino Americana e do Caribe em Ciencias da Saude (LILACS) database improves systematic reviews. Int J Epidemiol 2002;31(1):112-4.

Address for correspondence

Paula Otero, MD

Departamento de Informacion Hospitalaria

Hospital Italiano de Buenos Aires

Argentina

Email: paula.otero@hospitalitaliano.org.ar