Identifying a core set of medical informatics serials: an analysis using the MEDLINE database*

By Dean F. Sittig, Ph.D.
Assistant Professor of Biomedical Informatics and Biomedical Engineering

The Informatics Center
412 Eskind Biomedical Library
Vanderbilt University
Nashville, Tennessee 37232-8340

A study was undertaken to test the hypothesis that a core set of medical informatics serials could be identified by using standard bibliometric techniques. All journal articles indexed by the National Library of Medicine between 1990 and 1994 were included. Articles were identified by using the "MEDICAL INFORMATICS" Medical Subject Heading (MeSH) term. Each serial title containing articles was then ranked according to (1) the total number of medical informatics journal articles indexed and (2) the percentage of medical informatics journal articles indexed. Twenty-eight serials had more than 100 articles indexed under the "MEDICAL INFORMATICS" MeSH term. Thirty serials had more than 40% of their articles indexed under the "MEDICAL INFORMATICS" MeSH term. A "core" set of fourteen serials had 100 or more medical informatics articles indexed, including more than 70% of all articles they published. The methodology described provides librarians with another tool to use in the difficult task of journal selection. The set of "core" serials identified provides librarians with a ranked list of serials, based on which a medical informatics collection can be developed.

INTRODUCTION

Medical informatics is, by its very nature, multidisciplinary. Therefore, the question of how to define this emerging and diverse field often arises. Numerous experts in medical informatics have put forth definitions of the field [1–13]. The sheer number of definitions serves to prove that there is no generally accepted definition. A consequence of this failure to agree on a definition is that librarians and researchers alike have difficulty identifying the core serials of the field. At the same time, the National Library of Medicine (NLM) has taken steps to define the field through its use of Medical Subject Headings (MeSH) to index the scientific biomedical literature [14].

A study was undertaken to test the hypothesis that a core set of medical informatics serials could be identified by using standard bibliometric techniques. The primary objective was to identify the serials that publish the most or the highest percentage of medical informatics journal articles as defined by NLM through its indexing in the MEDLINE database for the years 1990 to 1994. A secondary objective was to evaluate the utility of two well-known bibliometric analysis techniques, Bradford's law and the Pratt Index, in the field of medical informatics.

* This work was supported in part by grants R29-LM0524 and G08-LM05443 from the National Library of Medicine.

BACKGROUND

MeSH

MeSH, one of the world's preeminent controlled vocabularies, was created and is maintained by NLM [15]. Most of the indexing terms under the "MEDICAL INFORMATICS" subject heading were added in 1987 following the work of Rada et al. [16]. One of the most important aspects of the MeSH vocabulary is its hierarchical structure, which allows terms to be linked to other terms that are broader or narrower in meaning. In addition, each term may be linked to multiple broader terms, allowing for a natural approach to vocabulary construction and article indexing.

The quality of an information-retrieval system is determined in large part by the selection of indexing terms and the assignment of those terms to the documents represented in the system. At NLM, experienced indexers analyze each article by reviewing the complete text, and then assign one or more MeSH terms—along with specific qualifiers, check tags, and publication types—to each article record, through the use of an automated system that helps to prevent errors [15]. Thus, the selection and assignment of terms is precise and highly consistent, which makes the information retrieval system extremely effective.

NLM defines medical informatics as "the field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine" [17]. This broad definition was used for the present study. The "MEDICAL INFORMATICS" MeSH term is divided into two main groups, "MEDICAL INFORMATICS APPLICATIONS" and MEDICAL INFORMATICS COMPUTING." Sample narrower terms within the applications grouping are "COMPUTER-ASSISTED DECISION MAKING," "INFORMATION STORAGE AND RETRIEVAL," and "INFORMATION SYSTEMS" [18]. Sample narrower terms within the computing group include "COMPUTER SYSTEMS," "COMPUTING METHODOLOGIES," "COMPUTING MILIEU," and "SOFTWARE."

Bradford's law

Bradford's law states that "[i]f scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as 1: n : $n2$: . . ." [19]. Bradford's law has been used extensively in the information science literature to describe the dispersion of articles in any scientific field [20, 21] and to identify a "nucleus" of serial titles [22].

Pratt's Index of Concentration

Pratt's Index permits comparisons of the dispersion of articles among different scientific fields. Pratt's Index of Concentration for rank-frequency distributions is defined as follows:

$$\text{Pratt's Index of Concentration } (C) = 2[(N + 1)/2 - q]/(N-1) \quad [23]$$

where N is the total number of items ranked, and q is the sum of rank times frequency divided by the total number of occurrences of all items.

Pratt showed that if all articles are concentrated in a single serial, then C=1.0; if all articles are distributed evenly among all serials, then C=0.0; and if they are distributed randomly according to a Whitworth distribution, then C=0.50 [23, 24].

STUDY METHODS

The CD-Plus version of the 1990–1994 MEDLINE database was used for all analyses. All articles categorized by NLM as journal articles (even if they came from the proceedings of a conference) that also were assigned the "MEDICAL INFORMATICS" subject heading were identified (i.e., the MeSH term was exploded, all subheadings were included, documents were restricted to focus, and the result was limited to journal articles). The source (serial title) for each article was then downloaded to a relational database. The number of medical informatics articles published by each serial were counted; this list then was sorted according to the number of articles published by each serial.

The total number of journal articles published and indexed in the MEDLINE database by each serial was counted. By using these figures, the percentage of articles classified under "MEDICAL INFORMATICS" was calculated for each serial. This list was then sorted according to the percentage of medical informatics articles published.

RESULTS

The CD-Plus MEDLINE file for 1990 to 1994 contains 1,537,320 journal articles indexed from 9,031 different serial titles. These titles represent approximately 4.5 years' worth of MEDLINE. There were 39,618 journal articles (2.6% of all articles indexed) indexed under the "MEDICAL INFORMATICS" MeSH term during this period. Of these journal articles, 19,895 (1.3% of all journal articles and 50.2% of all medical informatics articles) were indexed by NLM as primarily concerned with (i.e., restricted to focus on) medical informatics. Of the more than 9,000 serial titles, 903 published ten or more medical informatics journal articles, 2,474 serials published two or more, and 800 serials published a single medical informatics article during this period. When the medical informatics serials were "restricted to focus," it was found that 388 serials had published ten or more journal articles, 1,620 serials had published two or more, and 730 serials had published a single journal article concerned primarily with medical informatics during this period.

Table 1 lists the fourteen serials that published 100 or more medical informatics journal articles (restrict-

Table 1
Serial titles with most medical informatics articles*

Rank	Serial title
1	*Proceedings of the Annual Symposium on Computer Applications in Medical Care*
2	*Computer Applications in the Biosciences*
3	*Computer Methods & Programs in Biomedicine*
4	*International Journal of Bio-Medical Computing*
5	*Journal of Digital Imaging*
6	*Computers in Biology & Medicine*
7	*Computers in Nursing*
8	*Computers & Biomedical Research*
9	*Methods of Information in Medicine*
10	*M D Computing*
11	*Medical Informatics*
12	*Journal of Molecular Graphics*
13	*Journal of Medical Systems*
14	*International Journal of Clinical Monitoring & Computing*

* All serial titles that had 100 or more medical informatics journal articles and greater than 70% medical informatics articles are sorted in descending order according to the product of the total number of medical informatics journal articles and the percentage of medical informatics journal articles they published.

ed to focus), more than 70% of whose contents consisted of medical informatics articles during the period studied. These serial titles accounted for 14% of all the medical informatics journal articles published during this period. The calculations were repeated with all the journal articles indexed under the "MEDICAL INFORMATICS" term (i.e., not restricted to focus), and only the *Journal of Computer-Aided Molecular Design* was added to the list.

Table 2 presents the thirteen serials, more than 40% of whose articles were on medical informatics during the period studied, but which have published (or had indexed) fewer than 100 articles on the subject. These titles include up-and-coming serials, those no longer being indexed, and very small serials.

Table 3 presents the fourteen serials that published more than 100 medical informatics journal articles during the period studied, but fewer than 40% of whose articles concerned medical informatics.

Application of Bradford's law

The serial titles were sorted in rank order according to the total number of medical informatics journal articles indexed in MEDLINE (1990–1994). The cumulative sum of journal articles was plotted against the log of each journal's rank (i.e., a Bradford scatter plot) (see Figure 1). The slope (k=18433.4) of the linear portion of the curve was calculated for the line between the points that accounted for 33% (84 titles) and 83% (1008 titles) of the total number of journal articles (39,618) indexed. The x- and y-intercepts of this line were also calculated. The equation for the exponential portion of the curve was calculated by using

$$B=1/1n\,[R(c)/N]$$
$$B=0.715 \qquad [27]$$

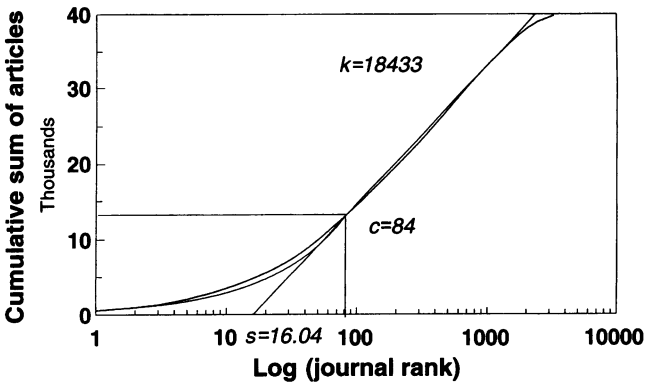
Table 2
Serial titles that published fewer than 100 medical informatics journal articles are sorted in descending order according to the percentage of medical informatics journal articles published and indexed in MEDLINE from 1990 to 1994.

Rank	Serial titles	% Medical informatics articles	Notes
1	*Artificial Intelligence in Medicine*	94.7	Indexing beginning 1993
2	*International Journal of Neural Systems*	94.3	Indexing beginning 1993
3	*Nursing Educators Microworld*	87.9	
4	*ISA Transactions*	75.0	1990 and 1992 only
5	*Journal of Chemical Information & Computer Sciences*	67.3	Selectively indexed
6	*Nurse Managers Bookshelf*	57.1	1990 only
7	*Kentucky Dental Journal*	57.1	
8	*Issues*	55.6	1990 only
9	*Medical Decision Making*	47.2	
10	*Ultrasonic Imaging*	44.2	
11	*Journal of the Society for Health Systems*	41.4	
12	*Medical Engineering & Physics*	41.2	
13	*Reflections*	40.7	Began 1992

Table 3
Serial titles that have published more than 100 medical informatics journal articles over the past 4.5 years, but have not published more than 70% medical informatics articles, are sorted in decreasing order according to the product of the total number of medical informatics journal articles and the percentage of medical informatics journal articles they published.

Rank	Serial titles	Medical informatics articles
1	*IEEE Transactions on Biomedical Engineering*	284
2	*Biomedizinische Technik*	209
3	*Computerized Medical Imaging & Graphics*	126
4	*Journal of Electrocardiology*	140
5	*Medical Physics*	191
6	*Magnetic Resonance in Medicine*	220
7	*Statistics in Medicine*	157
8	*Medical & Biological Engineering & Computing*	108
9	*International Journal of Radiation Oncology*	165
10	*Investigative Radiology*	121
11	*Pace: Pacing & Clinical Electrophysiology*	112
12	*Radiology*	146
13	*American Journal of Roentgenology*	123
14	*Nucleic Acids Research*	179

Figure 1
Bradford's law of scattering (*the medical informatics literature*)



where c is the inflection point between linear and exponential portions of the plot,

$R(c)$ is the cumulative sum of journal articles at point c, and
N is the total number of serial titles.

Therefore, the Bradford plot can be described mathematically by

$R(n)=557*n^{0.715}$ for $(1<n<c)$
$R(n)=18433.4*\log n/16.04$ for $(c<n<N)$

where 557 is the number of journal articles indexed from the serial with the greatest emphasis on medical informatics, *Proceedings of the Annual Symposium on Computer Applications in Medical Care.*

Application of Pratt's Index of Concentration

The "concentration" of medical informatics journal articles among the serial titles indexed in MEDLINE (1990–1994) was calculated to be 0.70.

DISCUSSION

A core set of medical informatics serials

The small "core" set of medical informatics serials (Table 1) represents the required holdings for any academic medical library interested in developing or maintaining a medical informatics collection. It is notable that no major biomedical resource library owns all of these titles. Ownership rates range from 5% of libraries for the least popular title to 79% for the most widely held title [26].

The serials identified in Table 2 that are still in publication or being indexed should also be considered important holdings because they represent the newer titles in the field. Finally, the serial titles that have published a large number of medical informatics journal articles, but not a high percentage (Table 3), should be considered for completeness.

It would be best to attach some absolute quality or indicator of importance to each journal article before performing the analysis described here. Unfortunately, such indicators are not currently available, although similar measures have been developed. For example, the Institute for Scientific Information in its *Science Citation Index* and *Journal Citation Reports,* calculates the "Impact Factor" and "Immediacy Index" [28] for thirteen medical informatics serials [26]. The present study simply attempted to identify the serials publishing the most or the highest percentage of medical informatics journal articles.

Bradford's law: is it applicable?

Bradford's law holds for the field of medical informatics, although it is unrealistic to expect every medical library to maintain eighty-four "core" serial titles. These eighty-four serials each had eighty or more journal articles indexed under the broad "MEDICAL INFORMATICS" MeSH term over the period studied. The next step was to look at the percentage of medical informatics journal articles published by each serial. Based on the percentage figures, the serial title ranked eighty-fourth would contain only 23.5% medical informatics journal articles. Thus, a library interested only in medical informatics articles would be forced to maintain a large number of mostly unrelated serials under Bradford's law. Therefore, for the present study, the total number of medical informatics articles was combined with the percentage of medical informatics articles to identify a "core" set of serials.

Pratt's Index

The medical informatics journal articles are concentrated to a greater extent than a random allotment would predict. The index of concentration (0.70) could be used as a basis for comparing medical informatics to other medical subspecialties, or to the larger general scientific fields.

CONCLUSION

This article presented a method of evaluating medical informatics serials that is based solely on the quantity and percentage of medical informatics journal articles published over a 4.5-year period. No attempt was made to judge the quality of any of the articles or the serials that published them. The methodology described provides the collection development librarian with another tool to use in the difficult task of journal selection. The resulting "core" set of medical informatics serials provides librarians with a ranked list

of serials based on which a medical informatics collection can be developed.

ACKNOWLEDGMENTS

Cheryl Chisnell, formerly of the Eskind Biomedical Library, provided expert technical assistance in the use of the CD-Plus MEDLINE database.

REFERENCES

1. BLOIS MS. What is medical informatics? West J Med 1986; 145(6):776-7.
2. COLLEN MF. Origins of medical informatics. West J Med 1986;145(6):778-85.
3. COOPER JA. Medical informatics: a definition and a role in academic medical centers. J Med Ed 1986;61(6):476-7.
4. LINDBERG DA, SCHOOLMAN HM. The National Library of Medicine and medical informatics. West J Med 1986;145(6): 786-90.
5. GREENES RA, SIEGEL ER. Characterization of an emerging field: approaches to defining the literature and disciplinary boundaries of medical informatics. In: Stead WW, ed. Proceedings of the Eleventh Annual Symposium on Computer Applications in Medical Care. Los Angeles: IEEE Press, 1987: 411-5.
6. HAUX R. On medical informatics. Meth Info Med 1989; 28(2):66-8.
7. STEAD WW. What is medical informatics? Colorado Med 1989;86(17):332.
8. LINCOLN TL. Medical informatics: the substantive discipline behind health care computer systems. Int J Bio-Med Comp 1990;26(1-2):73-92.
9. GREENES RA, SHORTLIFFE EH. Medical informatics: an emerging academic discipline and institutional priority. JAMA 1990;263(8):1114-20.
10. WARNER HR, BARNETT O, GREENES RA, CIMINO JJ. A view of medical informatics as an academic discipline [editorial]. Comp Biomed Res 1993;26(4):319-26.
11. SITTIG DF. Grand challenges in medical informatics? JAMIA 1994;1(5):412-3.
12. PROTTI DJ, VAN BEMMEL JH, GUNZENHUSER R, HAUX R, ET AL. Can health/medical informatics be regarded as a separate discipline? Meth Info Med 1994;33(3):318-26.
13. FRIEDMAN C. Where's the science in medical informatics? JAMIA 1995;2(1):65-7.
14. NATIONAL LIBRARY OF MEDICINE. Online indexing system fact sheet. Bethesda, MD: The Library, 1993 Apr.
15. NATIONAL LIBRARY OF MEDICINE. Medical subject headings (MeSH) factsheet. Bethesda, MD: The Library, 1993 Jan.
16. RADA R, BLUM B, CALHOUN E, MILI H, ET AL. A vocabulary for medical informatics. Comp Biomed Res 1987;20(3):244-63.
17. NATIONAL LIBRARY OF MEDICINE. MeSH scope note. In: Medical subject headings annotated database. Bethesda, MD: The Library, 1987 Jan.
18. NATIONAL LIBRARY OF MEDICINE. Medical subject headings: tree structures. Bethesda, MD: The Library, 1994.
19. BRADFORD SC. Documentation. London: Crosby Lockwood, 1948.
20. GOFFMAN W, WARREN KS. Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. Nature 1969;221:1205-7.
21. DROTT MC, GRIFFITH BC. An empirical examination of Bradford's law and the scattering of scientific literature. JASIS 1978;29:238-46.
22. GOFFMAN W, MORRIS TG. Bradford's law and library acquisitions. Nature 1970;226:922-23.
23. PRATT AD. A measure of class concentration in bibliometrics. JASIS 1977;28:285-92.
24. WHITWORTH WA. Choice and chance. New York: Hafner Publishing, 1948:205-8.
25. NATIONAL LIBRARY OF MEDICINE. SERLINE: explanation of contents of database. Bethesda, MD: The Library, 1993.
26. SITTIG DF, KAALAAS-SITTIG J. A quantitative ranking of the biomedical informatics serials. Meth Info Med 1995; 34(4):403-16.
27. BROOKES BC. Bradford's law and the bibliography of science. Nature 1969;224:953-6.
28. GARFIELD E, ED. Science citation index. Philadelphia, PA: ISI Press, 1992.

Received February 1994; accepted August 1995