

Appendix C.—MEDLARS Evaluations: A Review of the Literature

Introduction

This appendix reviews published and unpublished studies evaluating MEDLARS services and usage. In many ways, these studies have been dated by a decade of rapid technological advances in computerized retrieval systems. A paucity of studies is available on the system in place today; thus, generalizations based on the findings presented here may be misleading. The literature review here is offered to illustrate the kinds of questions one must ask in evaluating on-line retrieval systems, and the complexity and difficulty inherent in such an undertaking.

The National Library of Medicine (NLM or the Library) has devoted extensive time and other resources to internal and external review and has examined not only system performance but the basic purposes, goals, and constituency groups MEDLARS is designed to serve. Few institutions, public or private, have matched NLM's commitment to evaluation. This appendix reviews studies of user satisfaction, searcher variation, and the retrieval capabilities of the MEDLARS system. It also presents OTA'S evaluation of MEDLINE's coverage of selected topics. First, however, it describes the types, number, and motivations of MEDLARS users.

Users and Usage of MEDLARS

NLM collects utilization data for MEDLARS solely on an institutional basis. There is information on how many medical schools and hospitals have MEDLARS terminals, but there is relatively little information on the individuals requesting on-line searches, their reasons for seeking the information, or their level of satisfaction with search results. Further, and perhaps more importantly, there is no study available of the nonusers of MEDLARS and their reasons for not using its services.

Because so few data are available, users' expectations can only be appraised in light of their organizational affiliations, and the respective orientation of those institutions. In September 1980, 1,243 institutions in the United States had direct access to NLM's data bases. Forty-one percent of these institutional users were hospitals and clinics, 23 percent were commercial firms, and 9 percent were medical schools (see table 4 in ch. 2 for complete data).

This discussion defines the librarians and information specialists who conduct searches at computer terminals to be *intermediate users* of information systems, while *end-users* are the researchers and clinicians who

request that searches be done and who utilize the information retrieved. Sometimes, though rarely, end users conduct their own searches. In 1977, an estimated 80 percent of all searches were conducted by trained intermediate users alone, 20 percent by intermediate users with end users present at the terminal, and less than 1 percent by unassisted end users (160).

Patient Care Institutions

Hospitals are the largest and fastest growing group of institutional users of MEDLARS services. In 1978, hospitals represented one-third of NLM's on-line institutional users; by 1980, this figure had grown to over 40 percent. There are more than 7,000 hospitals in the United States; over 700 have MEDLARS terminals, and many others obtain on-line services from hospitals in nearby communities. In 1976-77, one study indicated that hospitals with terminals each processed an average of 493 MEDLINE searches per year (160). For that same period, professional schools conducted an average of 1,429 searches. The average for all institutions with access to MEDLINE was 842.

Hospital librarians report that about half their search requests are from physicians wanting information directly applicable to patient care. Often, the information is needed to aid the diagnosis and treatment of disease. In such instances, physicians are generally looking for a few relevant articles and need to obtain them within 24 hours. MEDLARS is also helpful to the physician preparing an article or lecture, or simply as a way of the physician keeping abreast of the literature on a particular subject or specialty. Nurses are also frequent users, often for reasons similar to physicians.' Questions on patient management represent the most common queries, though nurses also request searches for patient education, staff development programs, and the preparation of papers and presentations. MEDLARS is increasingly being used by ancillary service departments and administrative staffs in hospitals, especially for the planning of new services and the purchase of new equipment.

Though MEDLINE is the most frequently used data base in hospitals, the HEALTH and CANCERLIT files are also popular. Twenty-two percent of hospitals with on-line access to NLM's data bases also reported having used other non-NLM data bases in 1977, with PSYCHOLOGICAL ABSTRACTS and SCISEARCH the most often used. As the holdings of most hospital libraries are generally not extensive (the 1977 average number of serial subscriptions was under 400 for hospitals, but over 2,000 for professional schools) and

primarily clinically oriented, some document delivery problems are reported for articles identified through non-MEDLINE data bases. Sixty-one percent of all institutional MEDLARS users said they needed to acquire additional holdings based on the demand from on-line searches but did not have the funds to do so (26).

Research Institutions

The majority of institutional users are interested in MEDLARS as a research tool. These users include Government agencies, research foundations, professional associations, medical and professional schools, commercial research and development firms, and information brokerages. Little information is published on the utilization of MEDLARS services by such organizations, though some basic descriptive data has been reported by medical school libraries on their individual users (58,88,139,140,141). Although medical schools constitute only a small portion of research institution users, their experience with MEDLARS reflects that of similar institutions and is particularly useful for identifying the reasons that researchers request on-line searches.

MEDLARS usage in medical schools is characterized by utilization patterns similar to those found in hospitals. Both hospital and medical school libraries conduct searches for physicians, nurses, lab technicians, and administrators, and often for the same reasons—patient care, preparation of articles and lectures, and planning. Medical school libraries, however, see the majority of searches requested for research purposes (58,88,139,140,141).

While health practitioners tend to want a few relevant articles, researchers more often request broad, comprehensive searches. Faculty members, who divide their time between teaching, research, and patient care, tend to be the heaviest users of MEDLARS, requesting information needed for both ongoing and prospective research. For them, MEDLARS is an invaluable means of saving time and effort; for identifying articles that might be missed in a manual search; and for ensuring that they have a comprehensive bibliography available, especially when considering research in a new field (139,141).

Clinical Librarianship Programs

Clinical librarianship programs first appeared in the early 1970's, in response to the need for more timely dissemination of current developments and research findings. Primarily undertaken in large medical centers and teaching hospitals, the programs are designed to bring the skills of medical librarians directly to the

delivery of health services by providing "high, specific, case-related medical literature . . . in a manner that permits the information obtained to influence ongoing case management" (14). In a technical sense, the success of such programs depends on the librarian's ability to conduct highly specific, narrowly focused searches, in a brief period of time, ensuring that the relevant information reaches the clinician in time to influence the care of the patient. The ultimate objective of clinical librarianship is improved patient care.

Although these programs have received considerable attention and several articles are available describing their focus, methods, and results (129), evaluative efforts found in the literature have been limited in scope and have not adequately assessed any effects of such programs on patient care. They do report, however, significant increases in utilization of hospital library services resulting from the introduction of clinical librarians. For instance, one study showed a 120-percent increase in search requests from a staff with a clinical librarian program (61). A second study reported that 92 percent of physicians said that they read the articles given to them by clinical librarians, 86 percent said they learned something new from them, and another 20 percent indicated that the articles affected patient management to some degree (129). This study hypothesized that clinical librarianship programs could be a cost-effective alternative to many ancillary services, by substituting relatively inexpensive literature searches for more costly diagnostic tests.

MEDLARS Evaluation Studies

OTA reviewed three types of MEDLARS evaluation studies. The first group of studies, user satisfaction studies, asked end users to determine the relevance of retrieved documents to their requests. End users queried in these studies generally found MEDLARS search helpful in their research and/or clinical practice. A second group of studies evaluated variation in MEDLARS search results based on characteristics of intermediate users. These studies indicate that the training and experience of the searchers and the interaction between intermediate and end users in formulating search requests are key to the success of the on-line search. Finally, Lancaster's landmark study (77) evaluated the retrieval capabilities of MEDLARS before it was available on-line. Lancaster reported average recall and precision ratios of 58 percent and 50 percent respectively, found extraordinarily high variability from question to question searched, and conducted an extensive analysis of recall and precision failure. On the whole, Lancaster found little critically wrong with the system.

The fourth topic discussed below is OTA's evaluation of MEDLINE's coverage of selected topics.

User Satisfaction Surveys

Several librarians have attempted to determine how well MEDLARS satisfied the information needs of end users in their individual libraries. In 1977, Brown (22) reported that users of on-line services, including MEDLARS, were "generally satisfied" with those services. The population studied represented only a fraction of those who could potentially benefit from on-line searches. Underlining the significance of the latter point, 2 years later a MEDLINE feasibility study undertaken at the behest of the Northeastern Consortium for Health Information was based on the theory that "most potential users and supporters of MEDLINE within hospitals [were] unaware of its usefulness and application" (89). A concerted effort to publicize the system's services through the NLM network increased demand significantly. The author concludes that the 13 member hospitals participating in the study could generate sufficient usage to justify the costs of MEDLINE, at least on a shared-service basis.

Sharing access to NLM's data bases was also suggested in a study done on the use of on-line services in academic settings not affiliated with medical schools (11). Here, interest in such systems was identified in students and faculty members engaged in research in disciplines related to the health sciences. In each of these studies, users were given the opportunity to have searches conducted at minimal (or no) cost, and then asked whether they would request searches if charges for on-line time were somewhat higher. The majority of users in each study said that they would pay, although instituting charges was never tested. Nevertheless, by 1977, 90 percent of medical school libraries charged users for MEDLARS services (160).

Several evaluations have been conducted by submitting questionnaires to MEDLARS users, asking them to identify their purposes in requesting on-line searches and their satisfaction with search results. OTA reviewed two efforts undertaken in individual information centers (93,141), and two others that evaluated data combined from seven centers (139,140).

In one study, 246 of 428 users of MEDLINE services at the University of Virginia Medical Library responded to a survey asking whether searches had been of assistance to their research or clinical work, and whether they would continue to utilize the system after the imposition of modest charges. They were also asked if MEDLINE was a "substantial improvement over the traditional methods of searching through the printed indexers" (93). The study group was composed of nursing, medical, and graduate students, nursing

and medical faculty, and staff; health professionals outside the medical center were included in the study through a statewide medical information service. Over 93 percent indicated that MEDLINE had assisted in their research and clinical work, and that it was an improvement over manual searching. Seventy-five percent indicated that they would continue using the service after the imposition of charges.

The remainder of the studies reviewed in this section were conducted by Tagliacozzo of the University of Michigan, and are the most sophisticated user satisfaction studies of MEDLARS. In her first study, published in 1973, Tagliacozzo identified two sets of issues that must be considered in evaluating the service provided by any information system such as MEDLARS (141). The first relates to the performance and costs of the system. The second relates to the end users of the system: who they are, whether they represent all categories of users the service was designed for, and whether using the system altered the progress of their research and clinical practice.

The first Michigan study assumed that utilization of an information service could be explained, and to some extent predicted, on the basis of end-users' subjective assessments of its usefulness. Questionnaires were distributed to 275 MEDLARS users. Of 168 completed surveys, 7.2 percent reported their MEDLARS search as "not helpful," 25.9 percent as "moderately helpful," and 66.9 percent as either "helpful" or "very helpful." The respondents listed a variety of reasons for requesting on-line searches, though clearly the majority of requests was for ongoing or prospective research. Only 24 respondents classified their work as "exclusively clinical," a number far too small to draw any conclusions about the role of MEDLARS as a source of information for practitioners.

In 1975, Tagliacozzo published a second study, specific to MEDLINE, and drawing on data collected in seven Midwestern medical centers (139). The study examined the characteristics, motivations and purposes, expectations, and perceptions of MEDLINE users at a time when the system was just beginning to be used in the medical centers.

Tagliacozzo found that users tended to be either research faculty members working in the basic and clinical sciences, or students working toward advanced degrees or in clinical training. Again, it was difficult to determine the role of MEDLINE in the acquisition of knowledge, though it was apparent that use of the system was not confined to academicians. Most MEDLINE users reported that the retrieved information was primarily for research purposes and that the search was requested because the service was perceived "as a more effective means of reaching relevant citations than the traditional bibliographic instruments."

Tagliacozzo's third study allowed users to distinguish between the value of the search process and the usefulness of the search results, rather than simply offer an overall judgment of the service (140). Data were again collected by questionnaire in seven Midwestern medical centers. Sixty percent of the respondents reported positive reactions to search results, though many others noted that while the search had not provided many useful citations it had saved the time and energy required for a manual search, or had confirmed opinions that all relevant literature for a subject had previously been identified. In her conclusion, Tagliacozzo cautions against taking the users' judgment at face value, especially for determining whether information needs are satisfied through use of an on-line system, because so many factors, such as familiarity with relevant literature and the nature of the search requests, can influence users' responses to the system.

These studies leave the impression that end users find MEDLARS helpful to their work in research and clinical practice. However, for the most part they did not sufficiently test critical questions such as: were the information needs of the user satisfied? and did the system provide the user with all, or most, of the relevant literature which the data base contains?

Evaluations of Intermediate User Variation

Several studies have examined variation in MEDLARS search results based on characteristics of intermediate users. Two studies examined the results of searches conducted by end-users themselves, without the assistance of a trained librarian. One found that "nonlibrarian users" were quite capable of interacting with MEDLINE, as measured by the number of modifications to their search statements made at the terminal (118). A second study reported usage of two NLM data bases, MEDLINE and TOXLINE, and concluded that, when access to terminals was provided to pathologists and pharmacists, nonmediated use by researchers (i. e., without the assistance of a trained MEDLARS searcher) could be beneficial, if such users were given a "minimanual" describing MeSH vocabulary (130).

A 1978 study compared results of MEDLARS searches conducted through different software packages, one available from NLM, the other from a commercial vendor, Bibliographic Retrieval Services (125). The study reported that searches could be done equally well on either system, but that they differed significantly from a technical standpoint. The differences were reported to be important only to intermediate users and did not affect search results. It

was recommended that these variations "justify the dual availability of the files."

A recent evaluation, funded by NLM, examines the effect that the type of user training has on searching style and performance (156). The study considered 535 searches from 191 intermediate users, and found no statistically significant differences in searcher performance between intermediate users trained by NLM and those receiving MEDLARS training "informally." Performance measures for all searches were reported to be **23 percent** for recall and **67 percent** for precision.

Evaluation of MEDLARS' Retrieval Capabilities

The only intensive effort to determine MEDLARS' ability to retrieve relevant information efficiently was conducted in 1968 before the system was available on-line. This was a study of batch processing; today's system is very different. Lancaster's landmark study, *Evaluation of MEDLARS Demand Search Service (77)*, reported the precision and recall performance of the system for over 300 search requests. Users (practicing physicians and researchers) were asked to assess the relevance of articles retrieved through MEDLARS searches. An article was considered relevant if it had "value to the user in relation to the information need that prompted his request." On the average, Lancaster found MEDLARS to be operating at a **58 percent** recall level and **50 percent** precision level. He noted that these averages, though characteristic of retrieval systems, might be misleading, since results of individual searches were widely scattered.

Much of Lancaster's study was devoted to an analysis of "search failure:" that is, the reason why searches did not identify more of the relevant literature and why so many irrelevant articles were retrieved. The author reported that **25 percent** of the recall failures and **17 percent** of the precision failures could be attributed, in part, to a communication breakdown between the user and the system. He recommended that search request forms be redesigned to more accurately reflect the information need of the end user. Changes in indexing, indexing language, and approaches to searching strategy were also recommended.

OTA's Evaluation of MEDLINE's Coverage of Selected Topics

OTA evaluated MEDLINE's coverage of literature relevant to five selected topics in biomedical and health services research. The topics and results are displayed in table C-1. Review articles on each topic were identified, and their bibliographies provided a relevance base of pertinent documents against which MEDLINE's coverage could be measured.

Table C-1 .—MEDLINE's Coverage of Selected Topics

Topic	Number of relevant citations	Percent in MEDLINE	Percent recall ^a
Hepatitis and hepatoma.	161	69%	840/0
Hemoglobin genetics	385	81	94
CT scanners	166	86	82
Patient compliance	323	62	77
Health care delivery	423	20	71

^aNUMBER of items retrieved/number of items included in the data base.

SOURCE: Office of Technology Assessment.

MEDLINE's coverage is more than adequate for the biomedical topics, but less so for those related to health services research. This finding is not surprising, and does not necessarily reflect a deficiency in NLM's system. Documents related and pertinent to the health services topics appear in a wide variety of publications, including law, business and public administration journals, conference reports, and monographs. These publications are not normally indexed for MEDLINE, and it is unreasonable to expect NLM to cover such diverse information sources. Coverage for these topics may be much better in other NLM data bases, especially HEALTH.

Conclusions

Evaluating any information system's ability to satisfy the needs of its users is an enormously complex

task, for at the most fundamental level a system's performance is based on human behavioral patterns that do not readily lend themselves to description by simplistic facts and outcome measurements. Human behavior affects every stage of the indexing, search, and retrieval processes-effects that are not reflected in measurements of recall and precision, nor in most studies of user satisfaction. Evaluation procedures specific to information science are being developed, and the study of information needs and systems capabilities is expanding to include communications to other than scientists and researchers, the traditional users of information services. But more creative approaches to evaluation are still needed if information processing and dissemination are to improve (28).

The findings of evaluation studies OTA reviewed are not generalizable to the present MEDLARS systems, nor to its entire user community. Many studies were conducted on earlier versions of MEDLARS, in essence evaluating a system vastly different from that in place today. Others examined a user population or set of search requests too small to carry external validity. Still others paid insufficient attention to the intricacies of the task before them. In the absence of soundly developed evaluation methods applicable to large, complex systems like MEDLARS, the limited value of these studies, though frustrating, is not unexpected.