Critique of an Evaluation of Software for Searching MEDLINE

Howard L. Bleich, M.D., Beth Israel Hospital and the Center for Clinical Computing, Harvard Medical School, Boston, Massachusetts

In a previous study, investigators at McMaster University compared 14 ways to search the MEDLINE database and concluded that the method that cost the least (the National Library of Medicine's ELHILL program) yielded the highest proportion of relevant articles, whereas the method that cost the most (PaperChase) yielded the least. There are serious defects in the study design that invalidate the authors' conclusions.

On August 1, 1974, a computer terminal with a cathode-ray tube was installed in the library of Beth Israel Hospital in Boston. The terminal was connected to a program called PaperChase, which provides access to MEDLINE, the National Library of Medicine's database of references to the biomedical literature. The installation of the computer terminal was not announced, and those who used it did so without help; no one stood by to offer assistance, and no user manual was provided.

To the investigators' surprise, between 4 P.M. and midnight of the first day, 12 people signed on and conducted searches [1]. The next day, 39 people conducted 75 searches. Soon, lines began to form at the terminal and a second terminal was installed. At the end of the first year, 1229 persons had conducted 10,678 searches, during which they displayed 474,096 references and selected 111,564 for printing [2]. Beth Israel Hospital had become the largest reported user of the MEDLINE database in the world [1].

Once end-users at Beth Israel discovered that they could search the MEDLINE database themselves, word spread to other institutions. In 1981, Georgetown University developed "mini-MEDLINE," a user-friendly program "similar to the PaperChase concept." [3] Mini-MEDLINE, too, became heavily used. CompuServe made PaperChase available on its information service; for the first time end-users throughout the country could search the MEDLINE database. The National Library of Medicine (NLM) then made ELHILL (a search program designed for specially trained librarians) available to end-users. BRS developed Colleague, Dialog developed Knowledge Index, and by the early 1980s there were many routes to MEDLINE, each with a different combination of user-friendliness, quality and quantity of references retrieved, time consumption, and cost. Most routes were difficult to use, and elementary and advanced training courses for aspiring searchers became available [4]. "Gateway software" appeared, promising a more cordial user interface; among the new products were SciMate, SearchMaster, and InSearch. Grateful Med and CD-ROM products would come later.

The question of which route to MEDLINE is best for the clinician is important. The investment in equipment can be substantial, and if the system is difficult to learn, the cost of training can be greater still. If the system is difficult to use, the effectiveness of searching will be compromised. Investigators at McMaster University compared various routes used to search MEDLINE [5]. They reported that the route that cost the least (the NLM's ELHILL program) yielded the highest proportion of articles that were generally and directly relevant, whereas the route that cost the most (PaperChase) performed the worst. The authors concluded that "for those who anticipate searching MEDLINE frequently, it appears well worthwhile to learn to use the NLM MEDLINE system directly."

As the developers of PaperChase, my colleagues and I were puzzled by the McMaster study. The use of PaperChase had continued to increase substantially, both at Beth Israel Hospital and elsewhere. In 1989, when unlimited use of PaperChase became available at the University of Michigan, usage there quickly increased to levels much higher than ever reported for any competing service at any institution in the world [6]. The rate of growth in individual

0195-4210/92/$5.00 © 1994 AMIA, Inc.

subscriptions, transiently damped in the immediate aftermath of the McMaster study, soon returned to its previous level. Not only did PaperChase continue to thrive, but the gateway software that the authors of the McMaster study recommended as the best (SearchMaster) was soon taken off the market.

The findings of the McMaster study are difficult to reconcile with data on usage that we and others have collected [2,3,6]. The study is now dated, but it continues to be widely quoted (17 times in the past three years), and its authors continue to publish in the field [7]. I have examined the McMaster study and found defects in its design, which appear to invalidate the authors' conclusions.

BACKGROUND

When the study at McMaster University was done, access to the MEDLINE database was available through four groups in North America: the NLM, BRS, Dialog, and PaperChase. The computer program that the NLM uses (ELHILL) and the programs from BRS (including Colleague) and Dialog (including Knowledge Index) are all derived from a single program (STAIRS) developed by IBM [8]. These programs, all based on the same principles, use similar commands and perform similar functions. PaperChase is an entirely different program; it works in a totally different way.

The investigators at McMaster evaluated the performance of the various programs provided by the four groups. In the case of ELHILL, BRS, and Dialog, they repeated their measurements with gateway software designed to make the programs more user-friendly. Four computer programs, three of which could be used with more than one type of gateway software, yielded 14 combinations for use in searching the MEDLINE database.

EXPERIMENTAL DESIGN

The study was based on six clinical topics. Trained librarians translated each topic into a specific search strategy. Two librarians then entered these predetermined strategies into each of the 14 test systems. The references printed were compared with the results of a search of the *Index Medicus*, a computer search of MEDLINE run with the ELHILL program of the NLM, review articles, and citations of clinical experts gathered by reference librarians. All searches were run on an IBM PC, Mannesmann Tally printer, and Hayes Smartmodem 1200 B using Smartcom II software through the Datapac communications network.

The stated purpose of the study was to help the clinician or end-user identify the best route to the MEDLINE database. Early in the article the authors mention that "to date, little or no comparative data have been available to allow clinicians who want to do their own searching to make a rational choice among the various access routes to MEDLINE." Later the authors say that their data support recommendations "for clinicians who wish to do their own MEDLINE searching." Such statements, together with references to electronic searching "from the clinic, bedside, or operating suite," lead the reader to expect data that compare the results of searches formulated by clinicians. In fact, however, the searches were formulated by librarians. No data on clinicians or end-users were collected. Reference librarians analyzed the searches, determined the search strategies, and specified each keystroke; research librarians then typed in these keystrokes as instructed.

BIAS IN STUDY DESIGN

Bias in Expertise with Software

A question then arises: Did the reference librarians who specified each keystroke have equal, or nearly equal, expertise in each of the systems being compared. We wrote the authors and asked. One reference librarian had performed 9000 searches over a period of 3 years, mainly with ELHILL but also with the programs from Dialog and BRS. Another had had 11 years of experience with Dialog's programs plus 2 years of experience with ELHILL and the BRS programs. Another had worked part time for a year with ELHILL, BRS, and Dialog. As noted above, the ELHILL, BRS, and Dialog systems — all based on the same program — are similar to one another.

Whereas the librarians reported years of experience with ELHILL, BRS, and Dialog, they had almost no experience with PaperChase. Studies have demonstrated that when end-users with no previous training sit down at a PaperChase terminal for the first time, they find enough references of value to make them return in large numbers to use PaperChase again [1,2,6]. On the other hand, no one would anticipate that inexperienced users of PaperChase (or any other program) could search as efficiently as reference librarians using the systems with which they have had extensive training and years of experience.

In the McMaster study the librarians' lack of experience with PaperChase was evident in a number of ways. First, when they performed timing tests with ELHILL, BRS, and Dialog, they printed references and abstracts with a single command. But when they used PaperChase, they printed the items one at a time. PaperChase permits users to print an entire list with a single command, but the authors lacked the experience with PaperChase to know how to do it.

Second, the librarians interacted with ELHILL, BRS, and Dialog through commands, whereas they interacted with PaperChase in novice mode, through menus. PaperChase has a command mode to speed things up for the experienced user, but the librarians did not know how to use it.

Third, the searches were done with Smartcom II software set up to emulate a Televideo terminal. PaperChase supports Televideo emulation, but the librarians did not know how to use it.

Bias in Population Sampled

To examine whether data from trained reference librarians could reasonably be applied to clinicians, we obtained the search strategies used by the authors for each of the six clinical topics studied. For example, for the topic "Can sputum cytology be used to screen for lung cancer?" the search strategy was as follows:

(1) SPUTUM with subheading CYTOLOGY
(2) LUNG NEOPLASMS
(3) MASS SCREENING or COMPARATIVE STUDY or PROSPECTIVE STUDIES
(4) 1 and 2 and 3
(5) limit above set to ENGLISH language
(6) limit above set to HUMAN

This is an excellent strategy for a librarian to use with ELHILL, BRS, or Dialog. It is not a strategy that would occur to a clinician, and it is not a strategy that anyone should use with PaperChase. A clinician might look for SPUTUM, CYTOLOGY, and LUNG CANCER, but would probably not know that CYTOLOGY has to be identified as a subheading, or that in the MEDLINE database articles that deal with lung cancer are indexed under the medical subject heading LUNG NEOPLASMS.

In our experience most end-users are not familiar with the controlled vocabulary of 16,000 words and phrases chosen by the NLM as medical subject headings (MeSH terms). They are not familiar with the hierarchical tree structures of the MeSH terms; with their permutations, see alsos, and cross references; with which subheadings can be applied to which headings; with the appropriate use of explosions; with the standard abbreviations for authors' names and journal titles; or with the importance of spelling variants. Most end-users do not know about ENGLISH and HUMAN — entries that are superfluous with PaperChase. End-users confuse Boolean AND with Boolean OR, and they sometimes AND themselves into a null retrieval, even when one of the lists contains only a few postings [9,10]. Most end-users have never taken courses in electronic searching, or read books on how the MEDLINE database is put together. They need help to find the right MeSH terms, to broaden a search that is too narrow, and to narrow a search that is too broad. Alone among the systems tested, PaperChase guides the untutored user through each of these problems. By not testing the extent to which each system deals with the real problems of the untutored user [9,10], the authors introduced a major bias against PaperChase.

Bias in Search Strategies

The goal of each search in the study was to produce a printed list of apposite references. ELHILL, BRS, and Dialog are designed to produce such lists efficiently. That is the usual goal when a librarian performs a search on behalf of a client. But when a clinician performs a search, he or she is apt to want — not a list of references — but the answer to a specific medical question [11]. Indeed, it is believed that half of all searches of the National Library of Medicine's databases now relate to answering a specific medical question [12]. The clinician studies abstracts on line, looking for reports in thoughtfully reviewed journals that deal with the question at hand. Once the question has been answered, subsequent items are not displayed. Of the programs tested, only PaperChase displays references from the most highly regarded journals first. Failure to test which program could best answer a clinical question introduced another bias against PaperChase.

Bias in the Timing Tests

To search the literature with a computer, the user makes lists of references and manipulates these lists until the desired references are found. PaperChase permits the user to display each reference on a list as soon as it is found. The other three programs, in contrast, cannot display any of the references on a list until the computer has found them all. Since long lists

can take considerable time to complete, it is helpful to be able to display references while the computer continues its search. By comparing the time taken to complete each list, rather than the time taken to begin the display, the authors introduced yet another bias against PaperChase.

Bias in Definition of Gold Standard

The standard against which the authors compared the results of each search included a MEDLINE search with the ELHILL program. Not all MEDLINE programs retrieve the same references [13]. To the extent that one of the programs tested (in this case, ELHILL) contributed to the definition of the gold standard, one would expect that program to perform the best. For the comparison to be fair, all programs being tested should contribute equally (or not at all) to the gold standard.

Bias in Cost Data

The cost comparisons in the study did not include the cost of offline search preparation. When trained librarians search with ELHILL, they are advised to prepare for the search in advance using materials available from the National Technical Information Service. By contrast, when users search with PaperChase, there is no need to prepare in advance; the contents of these materials are built into the program. By excluding the cost of offline preparation for a search, the authors introduced a bias against PaperChase.

More important, by not knowing how to use PaperChase and by trying to make it perform like the programs with which the librarians were familiar, the authors introduced a substantial bias against PaperChase.

CONCLUSION

The question that the authors pose is, "Which route to the MEDLINE database is best for the clinician?" The question that their data answer is, "Which route to the MEDLINE database is best for the librarian with years of experience with ELHILL, some experience with the programs from BRS and Dialog, and virtually no experience with PaperChase?" The answer to this latter question is ELHILL, BRS, or Dialog, all of which perform about the same. (The data also indicate that experienced librarians need not bother with the various software packages designed to make ELHILL, BRS, and Dialog more user-friendly.)

It would be important to identify the best way for clinicians and end-users to search the MEDLINE database. There is a need to compare the costs and time spent by clinicians using ELHILL (with and without Grateful Med), PaperChase, BRS, Dialog, and the CD-ROM products, with each trying to answer the same medical question and each trying to produce a useful list of references. To be meaningful, such a study would need to enlist clinicians and end-users who had comparable amounts of experience with each of the programs being tested, and who were using equipment that was equally appropriate and properly configured for each system. To be meaningful, such a study would need to compare each of the programs against the same unbiased gold standard.

References

[1] Horowitz GL, Bleich HL. PaperChase: a computer program to search the medical literature. N Engl J Med 1981; 305: 924-30.

[2] Horowitz GL, Jackson JD, Bleich HL. PaperChase: self-service bibliographic retrieval. JAMA 1983; 250: 2494-9.

[3] Broering NC. The mini MEDLINE SYSTEM: a library-based end-user search system. Bull Med Libr Assoc 1985; 73: 138-45.

[4] Bredderman PJ, Wojcik J, Flake D. BRS Colleague MEDLINE end user training follow-up survey. Med Ref Serv Q 1988; 7(3): 15-34.

[5] Haynes RB, McKibbon KA, Walker CJ, et al. Computer searching of the medical literature: an evaluation of MEDLINE searching systems. Ann Intern Med 1985; 103: 812-16.

[6] Grefsheim SF, Shipman BL, Schwartz DG, Dow SC. Searching Medline at the University of Michigan. MD Comput 1991; 8: 32-6.

[7] Walker CJ, McKibbon KA, Haynes RB, Johnston ME. Performance appraisal of online MEDLINE access routes. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care. November 8-11, 1992. New York: McGraw Hill.

[8] Saton G. Introduction to modern information retrieval. New York: McGraw Hill, 1983.

[9] Walker CJ, McKibbon KA, Haynes RB, Ramsden MF. Problems encountered by clinical end users of MEDLINE and GRATEFUL MED. Bull Med Libr Assoc 1991; 79 :67-9.

[10] Kingsland LC, Harbourt AM, Syed EJ, Schuyler PL. Coach: applying UMLS knowledge sources in an expert searcher environment. Bull Med Libr Assoc 1993; 81: 178-83.

[11] Horowitz GL, Jackson JD, Bleich HL.

PaperChase: computerized bibliographic retrieval to answer clinical questions. Methods Inf Med 1983; 22: 183-8.

[12] Lindberg DA. Information systems to support medical practice and scientific discovery. Methods Inf Med 1989; 28: 202-6.

[13] Feinglos SJ. MEDLINE at BRS, DIALOG and NLM: is there a choice? Bull Med Libr Assoc 1983; 71: 6-12.