

## TERM AND CITATION RETRIEVAL: A FIELD STUDY

MIRANDA LEE PAO

School of Information and Library Studies, The University of Michigan

(Received 31 October 1991; accepted in final form 20 February 1992)

**Abstract**—The relative efficacy of searching by terms and by citations is investigated with real searches collected in health sciences libraries. The objective is to seek evidence to confirm or refute findings from a controlled pilot study, and to understand the factors at work in operational search environments. Overall confirmation was found. In both the pilot and field studies, the improvement of the odds that overlap items retrieved would be relevant or partially relevant was truly astounding. If an item was retrieved from both MEDLINE® and SCISEARCH®, it was six times more likely that it would be relevant or partially relevant as opposed to being not relevant, and 8.4 times more likely for definitely relevant retrievals. In the field setting, citation searching was able to add an average of 24% recall to traditional subject retrieval. Term or citation searching from the open literature produced lower precision results. Attempts to identify distinguishing characteristics in queries which might benefit most from additional citation searches proved to be inconclusive. In spite of the obvious gain shown by citation searching, online access of citation databases has been hampered by their relative high cost.

The reported work is an extension of a pilot study of the characteristics and retrieval effectiveness of two subject searching modes. The two search approaches available on commercial bibliographic databases are semantic retrieval based on text words, assigned keywords and descriptors, and pragmatic retrieval based on citations. The earlier study was an experiment performed on a data file with narrow subject focus (Pao & Worthen, 1989). The database was constructed such that its documents content was retrievable by descriptors and text words as well as by cited references. Thus, direct comparison of the retrieval results was possible from parallel searches using appropriate terms and citations on identical queries.

Citation searching in the control setting was found to add an average of 14% of relevant documents to a search. This confirms earlier indications from other researchers that citation searching complements searching by terms (McCain, 1989; Pao, 1986; Salton, 1971). A logical follow-up question is whether these results could be of practical use to the online searcher. Can the searcher expect similar results when searches are done on commercially available databases? For example, what is the percentage of MEDLINE® search topics which could benefit from a citation search? One also wonders how generalizable and how stable the findings are if a sample of real searches was collected from libraries where a wide variety of topics was searched. While only a few overlap items were found, these common documents tended to be highly relevant to the search topic. What are the odds that retrieved items derived from both types of search methods are relevant? Is a higher yield from a citation search related to specific types of topics? Obviously, knowledge of this type could be useful to the online searcher.

This second study is framed around the following aims:

1. to seek confirmation of the earlier findings by testing the two retrieval modes with real searches processed on commercially available databases and by evaluating search results by requestors who posed the queries;
2. to determine the magnitude of incremental contribution of citation retrieval to MEDLINE searching;

This work was supported by the NLM Grant RO1-LM-04177 from the National Library of Medicine.

3. to demonstrate the practicality of the combined use of keyword and citation searching; and
4. to suggest query characteristics which relate to each of the two search modes, such that citation retrieval for specific types of queries could be used more effectively by the online searcher.

## BACKGROUND

In general, it has been known for some time that citation relationships could be used to complement conventional subject searches in the retrieval of relevant documents (Salton, 1971). Recent research also draws similar conclusions that retrieval performance could be improved if both subject and citation searching were used (Chapman & Subramanyam, 1981; McCain, 1989; McCain, White, & Griffith, 1987; Pao, 1986; Shaw, 1990; White, Griffith, Cowen, Selinger, & Steere, 1984). Nevertheless, previous works comparing the relative retrieval effectiveness of citation and subject approaches were either based on small samples of searches or conducted on specially constructed document collections of limited size. These files were usually narrowly focused in terms of subjects. Furthermore, investigations on the use of citation as a retrieval strategy were often based on experimental algorithms unavailable to the average online searcher (Bichteler & Eaton, 1980; Chapman & Subramanyam, 1981; Pao & Fu, 1985; Shaw, 1990; White et al., 1984). A review of relevant studies has been included in the pilot study report (Pao & Worthen, 1989). They suggest that for a comprehensive search, all available search modes should be used. Only a few directly compared term and citation retrievals using the same set of queries (McCain, 1989). A much larger body of studies has shown that different fields of the bibliographic record also appear to access different documents even when they are on the same topic (Katzer, McGill, Tessier, Frakes, & DasGupta, 1982). As part of an elaborate and comprehensive study of online searching and searchers, Saracevic and Kantor (1988a) suggested that different "searchers tended to see different things in a question, and find different answers." Although citation searching was not specifically tested in their study, their overlapping sets from searching the same query by different searchers were small, and the relevance odds for these common items were much higher than those found by any one searcher. Thus, corroborating evidence that different subject access points and different searchers appear to tap different subsets of relevant documents is beginning to emerge.

A brief summary of the pilot study follows (Pao, 1986; Pao & Worthen, 1989). Through the courtesy of the Information Service Department of Norwich Eaton Pharmaceuticals, Inc., Norwich, NY, a machine-readable file on a narrow subject, enteral hyperalimentation, was provided to the study. This was a working database created in-house to help researchers and professional personnel to answer questions on nutritional problems. Each item was abstracted and indexed in depth by the in-house staff (Worthen, 1979). The file of 1331 items with published works from 1956 to 1983 was culled from many indexes and journals. It was searchable by keywords in the title, abstract, and descriptor fields. A complementary citation file was created by the research team by linking items cited among the 1331 publications so that citation searching was also possible. During a six-month period, 33 queries searched by an experienced searcher at the company were collected. A registered dietician at Norwich Eaton supplied a seed document for each of the 33 topics. As a frequent user of the database, she was responsible for answering questions relating to diets and use of elemental diet products. In addition to the regular subject-based search, a citation search using the supplied seed document was also conducted for each query. Two subject experts as surrogates for the requestors evaluated the items retrieved. Results showed that, aside from confirming that term and citation retrieval complemented each other, citation searching was found to contribute an additional 14% relevant items to each search as a whole. Although only a few documents were retrieved by both methods, most of the overlap items were highly relevant. The obvious practical implication is that more relevant items on a given query could be retrieved by taking the union of term and citation searches, and a few relevant items could be expected from the intersection of the two retrieved sets.

Directly comparing the retrieval performance of subject terms and citation links, McCain (1989) analyzed the search results on nine topics. In addition to using three subject databases on DIALOG®, an average of 4.9 seed documents were searched on two citation indexes. Average relative recalls of 67% and 43% were found by using only descriptor-based searches in the subject files and only citation searches, respectively. These comprehensive searches produced an impressive average of 33% incremental relevant items from citation retrieval. There was a 10% overlap between the two search approaches. The author's pilot study produced similar results, namely, citation searching added to the relevant pool retrieved, and few items were common to both outcomes.

The complexity of subject retrieval has been called into focus in recent literature (Bates, 1986; Borgman, 1986). In the search for solutions, Bates has called for new system designs in which as many options as possible are made available and made obvious to the searcher (Bates, 1989). She cited six commonly used literature-seeking tactics. Clearly, more knowledge on the effectiveness of each tactic can be of immediate practical utility. Although citation searching promises to augment subject searching, the magnitude of the contribution of citation retrieval is inconclusive.

The purpose of the present study is to extend the investigation of citation and term retrieval to actual field settings. Although controls introduced in a specially built database and the relative homogeneity of search topics would be sacrificed, the significant gain of a field study is that the collected data would reflect the conditions typical of the day-to-day use of online systems. Confirming data is sought by examining a sample of real searches from a wide variety of biomedical topics. This study differs from the pilot study in that data were collected from several health sciences libraries in the course of their normal online searching services. In fact, normal charge fees were applied to each MEDLINE search, although each was accompanied by a free citation search. Only user-initiated searches were used, thus assuring that the queries represented real information needs. The retrieved items were evaluated for relevance by the user who originated the search. Only commercially available databases were used by qualified search professionals. Since the two types of subject access modes can only be conducted on two separate databases, any net gain should also be weighed against cost and practicality. The range of search topics also offered the opportunity to detect, if not identify, the types of biomedical queries which might benefit most from additional citation searches.

### *Methodological considerations*

Up to now, large-scale equitable comparisons of term and citation retrieval have been lacking. This is due in part to the way databases have been structured. Subject-based files do not contain citing data, and citation indexes do not include assigned descriptors and other content-bearing fields of the bibliographic record. In other words, upon logging on to a database, one is unable to pool the results of a descriptor-based search with one based on citation searching. The citation indexes from the Institute for Scientific Information are the only databases available for citation searching, and author-assigned index terms were introduced as a search field only a few months ago. To study the relative retrieval effectiveness of citation and term searching, one would have to search two different databases. In biomedicine, the coverage of the MEDLINE database is very different from the highly multidisciplinary SCISEARCH®, the online version of *Science Citation Index*®. To make a fair comparison, only those records covered by both databases must be used. This would mean the capture of a segment of a citation file and a subset of a descriptor-based file containing only records found in the citation file segment. If one planned to compare conventional subject searching with retrieval by co-citations or bibliographic coupling methods, one would have to prepare the citation retrieval data by various postsearch processing techniques, since bibliographic coupling and co-citation clusters are not directly searchable online. At present, such studies are of limited practical use for the searcher.

The above underscores the difficulties encountered in producing generalizable data on citation and term searching of the same query. Furthermore, the assessment of retrieval effectiveness necessitates relevance judgement of the retrieved items. Anyone who has col-

lected evaluated searches can attest to the difficulties in obtaining cooperation from users. Obtaining an adequate sample of real searches which have been evaluated by requestors with real information needs presents a considerable challenge to researchers. There is also the instability associated with the process of relevance evaluation itself. Nevertheless, it is easy to understand why it has been recommended that retrieval tests should be based on authenticated searches as much as possible (Sparck Jones, 1988). The validity of scaling up from small samples has been questioned by several researchers (Sparck Jones, 1981; Sparck Jones & Van Rijsbergen, 1976). Extrapolating from small samples presents unresolved problems. More recently, a similar criticism was voiced by Blair (Blair & Maron, 1990). Conflicting recall and precision results from different studies may have been derived from widely different collection sizes and small search samples. Sparck Jones urged that real databases be used for retrieval experiments. In spite of the obvious problems associated with field studies, there are many significant advantages.

## METHODOLOGY

### *Data sources*

Four health sciences libraries participated in collecting search data for the study. Each library serves a teaching hospital. A total of 102 mediated searches were performed on behalf of the study. There were 13 searches with no evaluation even though each requestor was contacted through follow-up letters. Two-thirds of the searches were collected from the Taubman Medical Library at The University of Michigan. Most requestors were faculty, staff, and house officers from the Medical School. A few others were associated with the other professional colleges in the health sciences.

In all cases, when a mediated search was requested on MEDLINE, the user was asked if he or she would be interested in receiving additional items on the same topic from a citation search and if the individual was willing to evaluate the search results for the use of the study. When cooperation was obtained, the patron was asked to supply information on one or two papers reflecting the content of the topic to be searched. These were used as seed documents for the parallel citation searches. The project then supplied the user with a free citation search. However, 90% of the queries were supplied with only one seed document. In two of the four libraries, the librarian personally assisted in suggesting the use of a citation search as a way to increase the recall of the requested topic whenever possible. To collect the necessary data from The University of Michigan, a modest campaign was launched to promote a free citation search to complement any regular search request on MEDLINE on the same topic. An advertisement was posted in the hospital bulletin to encourage cooperation with the "free" citation search as an inducement.

For each query, the online searcher searched both MEDLINE and SCISEARCH. Both printouts were presented to the patron. As in any realistic setting, a variety of requirements were being met by these searches. No limit on the scope of each search was imposed by the project, except those dictated by the requestor. As expected, the size of the retrieved sets as well as the relevant subsets varied widely among searches and topics. A couple of searches retrieved a grand total of two records each, and 472 unique items were retrieved for another topic. Similar to earlier field studies, most patrons requested high precision searches. All searchers participating in this project have many years of experience, although no attempt was made to control for the variations which are known to be introduced by the use of different searchers.

### *Authenticated searches*

Several criteria guided the study design. First, as mentioned above, all data were taken from *actual search requests* submitted to health sciences libraries so that a wide range of biomedical and other health-related topics was included. Most of the 89 queries related to clinical practice and biomedical sciences. There were also queries on behavioral problems associated with pregnancy and child development. One user asked to search information retrieval problems in postgraduate medical education. Topics ranged from those relating

to family medicine and health insurance issues to a highly specific query on esophageal polyps and Castleman's disease.

Second, searching was done on *commercially available databases*, namely, MEDLINE and SCISEARCH. Topics of citing papers collected for this study were limited only by the coverage of the citation database. Unlike the study on the subject of enteral hyperalimentation, these papers were not necessarily indexed directly under the topic searched. For example, if papers in cognitive psychology journals cited a highly relevant paper on Castleman's disease, they would be included as long as those journals were indexed by both databases. Most of the users supplied only one seed document which was used for the citation search.

Third, in spite of the instability associated with human relevance judgements, items were evaluated by the user who requested the search. This procedure approximated *realistic judgement* made in information encounters. The patrons were presented with two copies of both MEDLINE and SCISEARCH search results, one of which was theirs to keep. On the second copy, they were instructed to evaluate each printed citation by assigning one of the three scores:

- A for "relevant" (R) to the query
- B for "partially relevant" (pR) to the query
- C for "not relevant" (NR) to the query

In one library, hard copies were supplied free of charge. In that setting, only those items for which hard copies were requested were assumed to be relevant (R). In the final tally, items judged as "relevant" (R) or "partially relevant" (pR) were collapsed into a single "all relevant" group. Wherever appropriate, a "definitely relevant" group was also considered which consisted only of those items marked as "relevant" (R). Relative recall ratios were computed based on the "union of output," namely, all unique relevant or partially relevant items retrieved from the two parallel searches. These procedures were similar to those used in the pilot study.

#### *A common pool of records from MEDLINE and SCISEARCH*

Although both SCISEARCH and MEDLINE were used, the analysis was restricted to only those records which resided in both databases. In theory, a common journal pool should contain publications covered by both databases. Yet, for a given query, a set of records may be retrieved from one database and some of these may not be derived from the other file. However, for equitable comparison, analysis must be based on all papers as long as they have been indexed by both databases. Items common to both databases, whether they were retrieved or not, were established in three steps:

1. A list of common journals was compiled based on the published journal lists from the two database producers. This list permits the identification of their common journal coverage.
2. From the downloaded search results, the inconsistencies and errors in the journal title field had to be unified. This process of normalization enabled the identification of duplicate records from each pair of parallel searches. A substantial number of errors in the retrieved sets was found.
3. Even though a common pool of journals was identified, the indexing policy of the National Library of Medicine and the Institute of Scientific Information may differ in the number and types of paper indexed from these common journals. Since the stated policy of SCISEARCH is to index "from cover to cover" all journals in their journal list, any item retrieved from MEDLINE which has been published in the common journal set *must* be indexed by SCISEARCH. MEDLINE selectively indexes a group of journals in addition to the exhaustive indexing of a large number of core journals. Hence, if an item was retrieved from SCISEARCH and not retrieved from MEDLINE, then there was no assurance that it had been indexed by MEDLINE. Therefore, after the identification of the common journal set, a ran-

dom selection of papers retrieved from SCISEARCH only was searched by the technique of known-item search in the MEDLINE database. The purpose was to estimate the percentage of papers in the common pool of journals not covered by MEDLINE.

4. Additionally, since SCISEARCH started publication in 1974, all pre-1974 retrievals were excluded from data analysis.

A random selection of 10% of the items retrieved only from SCISEARCH was checked. It was found that less than 3% of the papers in the citation file were not indexed in MEDLINE. As shown in the following section, citation retrieved only one-fourth of the total 5558 unique items. Thus, the maximum number of items retrieved which might not have been contained in both databases was quite small. Hence, no attempt was made to identify and eliminate these papers.

## RESULTS

A total of 7208 records including duplicate items were retrieved by the 89 pairs of searches from the four libraries. They were found in 1588 journal titles, of which 915 were in the common pool of journals. Eliminating those papers published outside of the common journal pool, 6114 records were retrieved of which 5836 were unique items. Only less than 4.8% of the unique items, whether they were relevant or not, were retrieved by both search methods, a result similar to the 4.0% found in the pilot study. For the project, a total of 7208 separate relevance judgements were made.

### *Recall and precision*

All recall ratios computed are relative recall values. Overall, findings from the first study were confirmed. MEDLINE produced an average relative recall of 77% and precision of 56%, and SCISEARCH produced 33% recall and 60% precision when the "all relevant" group was used as the basis for analysis. The hypothesis that recall from MEDLINE is the same as that from SCISEARCH is rejected with a probability of  $p = 0.000$ . Figure 1 shows that, as expected, MEDLINE produced the majority of relevant retrieval for biomedical topics. If only one database is used, 68% of all MEDLINE searches achieved 70+% recall, while 66% of the citation searches fall below 40% recall. When only the definitely relevant documents were considered, a preponderance of 84% of MEDLINE

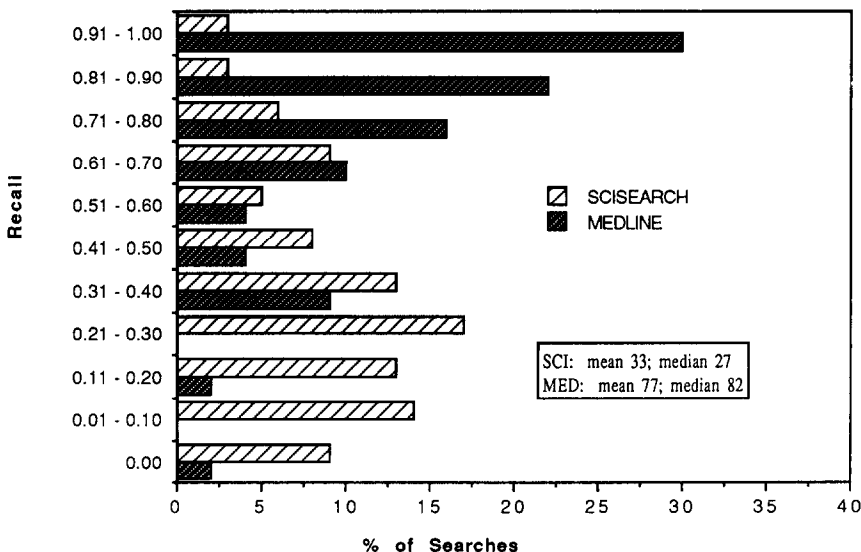


Fig. 1. Recalls of all relevant retrievals from MEDLINE® and SCISEARCH®.

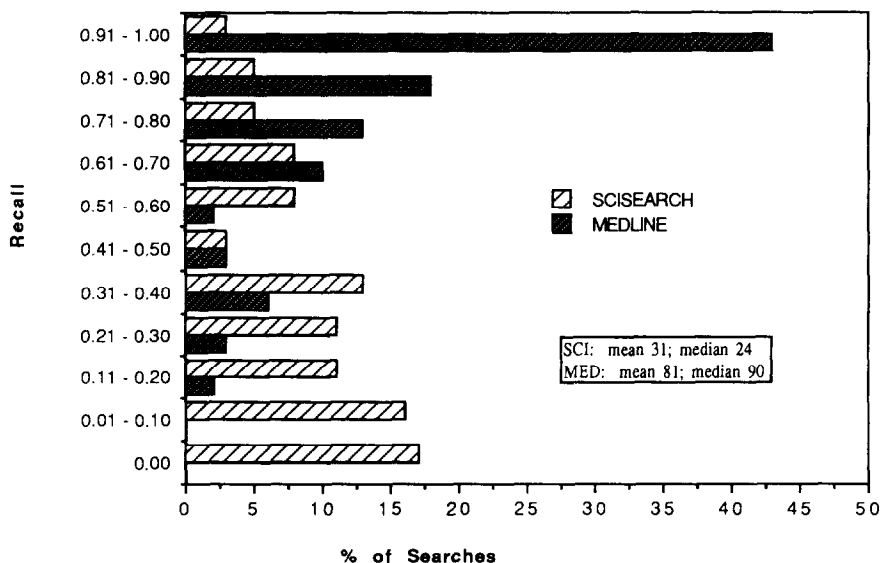


Fig. 2. Recalls of items of definite relevance from MEDLINE® and SCISEARCH®.

searches in Fig. 2 were shown above 60% recall, and 68% of those from citation were below 40% recall.

Table 1 compares the average recall and precision values obtained from the field and pilot studies. Interpretation of the recall and precision results must take into consideration that none of data sets are normally distributed, a finding emphasized by the Saracevic/Kantor study. The mean is often greatly inflated by one or two searches with high values. Both the mean and the median should be taken in consideration for a more accurate representation of the sample value. Since retrieval done in the pilot study was conducted on a subject file with up to 36 descriptors for each document, higher recall and precision from term retrieval were found. Overall, the field study did produce lower recall and precision by the subject approach. In particular, the median precision ratios for the field and pilot studies were 53% and 70%, respectively. In the open literature, one can expect to scan through many nonrelevant items. Yet, every MEDLINE search retrieved some items, and only one search yielded no relevant publications.

On the other hand, from the wide array of search topics, 5% of all searches produced no hit from SCISEARCH. Table 1 indicates 33% and 26% average relative recall ratios for the field and pilot studies, respectively, with accompanying 60% and 85% precision values from citation searching. The median recall was 27%, a significant increase from the 18% computed for the pilot study. Although there are more relevant items extracted from

Table 1. Performance measures for term and citation retrieval from pilot and field studies

	Recall					Precision				
	Mean	SD	Min	Max	Med	Mean	SD	Min	Max	Med
Field study (all relevant):										
MEDLINE®	0.77	0.23	0	1	0.82	0.56	0.26	0	1	0.53
SCISEARCH®	0.33	0.28	0	1	0.27	0.60	0.32	0	1	0.70
Field study (definitely relevant):										
MEDLINE	0.81	0.22	0.14	1	0.90	0.37	0.24	0	1	0.31
SCISEARCH	0.31	0.30	0	1	0.24	0.38	0.28	0	1	0.40
Pilot study:										
Terms	0.86	0.26	0	1	0.99	0.67	0.25	0	1	0.71
Citations	0.26	0.29	0.01	1	0.18	0.85	0.28	0.22	1	1.00

citation links, the trade-off is a considerable loss of precision. The median precision value is 70%, as opposed to 100% from the pilot study. This may be explained in part by the fact that this field draws upon a multitude of topics, including gastroenterology, surgery, and nursing, but citations for documents in the proprietary file only consisted of items directly indexed as topics in enteral hyperalimentation. In fact, only 509 of the 1331 papers were cited. On the other hand, citation searching in the open literature has the potential of retrieving any citing paper indexed under any topic as long as it is published in the common pool of journals. As shown repeatedly by the application of Bradford's Law on various subject literatures, coverage of any journal is seldom limited to a single topic. SCISEARCH, being a multidisciplinary database, permits a large number of potentially related subject links. The pool of citing papers extended well beyond the topic sought. Searching the open-ended literature in SCISEARCH appeared to increase recall, but it also tended to produce lower precision. Therefore, one could expect the bulk of relevant items for biomedical subjects to come from searching MEDLINE, but there were enough potentially useful items to make a SCISEARCH search worthwhile.

### *Relevance odds and search modes*

The effectiveness of searches as a whole is usually assessed by macro analyses using a search as the basis for analysis. The results are often expressed by recall and precision, and these are not the most sensitive measures. Particularly, when the two measures are not strongly correlated with other variables, Saracevic and Kantor (1988a) charged that their explanatory power is severely limited. They extended their analysis to the micro level which concentrated on the relevance of each item retrieved by a search. They were able to show many more significant relations among variables by the item-wise level of analysis. Using user's relevance judgement of each retrieved item as the basis for analysis, the micro level analysis has turned out to be much more powerful than the search-wise analysis. When appropriately used, the item-wise analysis seeks to explain the impact of a given variable on the odds that retrieved items were relevant, as opposed to not relevant. To do so, the two researchers applied the log cross-product ratio analysis to both the macro and micro analyses on their retrieval data. The cross-product ratio is also known as the odds ratio and is used frequently in analysis in biomedicine. Detailed accounts of the technique can be found in their three-paper series (Saracevic & Kantor, 1988a, 1988b; Saracevic, Kantor, Chamis, & Trivison, 1988).

The cross-product ratio analysis was used to determine the relevance odds of items retrieved from MEDLINE if they were not retrieved from SCISEARCH for the same topic. Two variables were identified; each must contain only two values. As a result, a  $2 \times 2$  table was set up. The independent variable was the search mode, with term searching and citation searching as the only two values. They were represented by MEDLINE and SCISEARCH. The dependent variable was the relevance judgement made by users on each item retrieved. The "all relevant" (R + pR) value is juxtaposed with the "not relevant" (N) category. Each item must fit into only one of the four resulting categories.

As only 4.8% of the items were retrieved by both search modes, they were excluded in the following analysis and analyzed separately in the next section. The answer to the following question was sought: *What were the odds that retrieved items would be relevant or partially relevant, as opposed to not relevant if they were retrieved only from term searching as opposed to only from citation searching and not from both?* Table 2 shows that there were 5558 items retrieved by one search mode alone. There were 4092 items retrieved only by MEDLINE, 2277 of which were judged relevant or partially relevant, while 1815 were judged not relevant. Similarly, of the 1466 retrieved by citations alone, 643 were judged relevant or partially relevant and 823 not relevant. The relevance odds were computed as follows:

Searching MEDLINE:  $2277/1815 = 1.26$

Searching SCISEARCH:  $643/823 = 0.78$

Ratio:  $1.26/0.78 = 1.61$



Table 2. Data for the calculation of relevant odds for items retrieved from each search mode—pilot and field studies

Relevance judgement	Search mode		Total (N)
	Terms only (N)	Citations only (N)	
<i>Field study</i>			
R + pR	2277	643	2920
NR	1815	823	2638
Total	4092	1466	5558
<i>Pilot study</i>			
R + pR	781	54	835
NR	399	150	549
Total	1180	204	1384

R: Relevant.  
pR: Partially relevant.  
NR: Not relevant.

Ln Ratio: 0.48

Standard Error (SE): 0.061

*t* value: 7.77

Since the absolute value of *t* exceeds 2, this result was statistically significant at the 0.05 level. It indicates that the relevance odds for an item increased by a factor of 1.61 if it was retrievable from MEDLINE only. In other words, the odds for an item to be relevant or partially relevant as opposed to not relevant were increased by 1.61 (or 61%) if it was found only in the MEDLINE output, as opposed to SCISEARCH only.

Data from the pilot study were also reanalyzed by the cross-product ratio technique (see Table 2). The relevance odds were computed as:

Term searching:  $781/399 = 1.96$

Citation Searching:  $54/150 = 0.36$

Ratio:  $1.96/0.36 = 5.44$

Ln Ratio: 1.69

SE: 0.155

*t* value: 10.90

This was also a statistically significant result at the 0.05 level. The odds that an item was relevant or partially relevant as opposed to not relevant were improved dramatically, by 5.44 times, for those retrieved by terms only, as opposed to those retrieved by cited references. An item retrieved from a closed file with a narrow focus had extraordinary high odds of being relevant if items were accessed by terms. This finding corroborates with the intuitive notion that this carefully compiled file with in-depth subject indexing could facilitate subject retrieval. On the other hand, since its citation links were severely limited, its citation retrieval capability suffered.

#### *Unique contribution by citation searching*

As mentioned earlier, 5% of the queries produced no retrieval from the citation index. Of the remaining searches, 89% did retrieve additional relevant items. These would be missed if searching was limited to using MEDLINE alone. A total of 1466 unique additional papers were retrieved from citation relationships. Of these, 326 items were judged relevant, 317 judged partially relevant, and 823 not relevant (see Table 2). In the field set-

Table 3. Distribution of recall from incremental citation retrieval

Recall	All relevant (% of searches)	Definitely relevant (% of searches)
0.00	10	25
0.01-0.10	20	23
0.11-0.20	25	11
0.21-0.30	14	13
0.31-0.40	11	11
0.41-0.50	5	2
0.51-0.60	8	6
0.61-0.70	3	3
0.71-0.80	0	3
0.81-0.90	2	2
0.91-1.00	2	0

All relevant recall: mean-0.24, median-0.19.

Precision: mean-0.56, median-0.63.

Definite relevant recall: mean-0.20, median-0.12.

Precision: mean-0.30, median-0.34.

ting, citation retrieval has added an overall average of 24% relevant papers per query (see Table 3). Nearly 44% of the retrieved items belonged to the "all relevant" group. Indeed, the additional 24% relevant materials from citation is higher than the 14% found in the pilot study, but not nearly as high as McCain's 33% figure (McCain, 1989). The skewed distribution of unique contribution by citation is displayed in Table 3. The table reads as follows: 20% of citation searches added 1 to 10% relevant items to each topic. If only definitely relevant items were considered, citation searching added 1 to 10% relevant retrieval to 23% of all searches. Cumulating the second through the fourth rows, 59% netted a gain of up to 30% relevant or partially relevant materials per search; 47% realized an increase of up to 30% additional items of definite relevance. While term searching could retrieve items from other related subjects, by its very nature, citation retrieval was able to link an overall average of 24% relevant sources not semantically related to the topic in question.

Citation searching added fewer than 16 items to two-thirds, or 66% of all topics in Table 4. Most searches added only a few items, and as the number of additional items from citation searching increased, a gradual rise of recall also followed. A correlation coefficient of 0.58 was computed for the number of added items and its gain in relative recall.

Table 4 also shows that although a citation search adds an average of 24% relevant materials, the average precision figures were in the 50s and 60s for those searches with less than 30 unique contributions, and the median was 63%. In other words, if only a few items were added from citation searching, at least half of them were likely to be relevant to the topic sought. Indeed, 83% of the queries had at least one item of definite relevance from citation links. These would have been missed if the citation database had not been con-

Table 4. Distribution of incremental contribution by citation searching

Retrieved documents (N)	% search	All relevant		Definitely relevant	
		Recall	Precision	Recall	Precision
1- 5	38	0.12	0.57	0.10	0.31
6-10	11	0.29	0.64	0.26	0.36
11-15	17	0.22	0.56	0.22	0.31
16-20	9	0.20	0.63	0.12	0.33
21-25	5	0.37	0.67	0.34	0.42
26-30	3	0.42	0.70	0.29	0.27
31+	17	0.48	0.36	0.30	0.17

sulted. While the median number of additional retrieved items was 10, 5 could be judged relevant.

In sum, in scaling up the experiment by the use of two real databases, proportionally more additional relevant retrievals from citation linkages were found. Although the majority of searches can be expected to derive only a handful of additional papers from SCISEARCH, over half of them are likely to be relevant.

#### *Overlap retrieval from MEDLINE and SCISEARCH*

An overlap or intersection set contains items retrieved both from MEDLINE and SCISEARCH. As reported in the pilot study and by other investigators, in spite of the extensive searching experience of the searchers, very few papers were retrieved by the intersection of these paired searches. From a total of 5836 unique items retrieved, only 4.8% (278) overlap retrievals were found. They included relevant and nonrelevant items. Twenty-eight percent of all search topics produced no overlap. Of the remaining 72%, 78% (relevant or not relevant) had less than seven overlap items, and 42% had only one to two common items (see Table 5). There were from 1 to 60 common items in each set, and an average of six common items per query was found.

Examining the *relevant* overlap retrieval, of the 278 overlap items, 197 (71%) were judged "relevant" (R) and 240 (86%) were "relevant" or "partially relevant" (R + pR). Although the 278 items constituted only 4.8% of the total retrieval for 89 topics, the distribution of relevant items was uneven. Excluding the 28% of searches with no overlap, 92% had at least one relevant or partially relevant overlap. The average overlap set contained 13% of relevant or partially relevant materials with 81% precision (see Table 6). If only definitely relevant items were tallied as relevant items, an average of 17% recall and 69% precision was found. All except one of the topics with overlap had at least one item of definite relevance.

Average recall and precision of the overlap sets in the pilot study were 12% and 92%, respectively. For the definitely relevant items, recall and precision of an average overlap for this specially constructed file were 16% and 75%, respectively. These results are very comparable between the two studies.

The "all relevant" columns in Table 5 show that 38% of the searches have one to two relevant overlaps. Sixty-three percent had one to four relevant items. Although 10% of those searches with overlaps had no definitely relevant items, 65% had one to four items of definite relevance in their overlap sets.

Furthermore, Fig. 3 gives the breakdown of the distribution of search topics with overlap items in terms of precision values. For an overwhelming 63% of the topics, all overlap items were relevant. At least half of the overlap items in 88% of topics were relevant.

Table 5. Distribution of the relevant overlap retrieval

Documents (N)	Total retrieval <sup>a</sup> (% searches)	All relevant (% searches)	Definitely relevant (% searches)
0		8	10
1- 2	42	38	48
3- 4	25	25	17
5- 6	11	10	8
7- 8	2	4	2
9-10	6	0	4
11-12	4	6	4
13-14	0	2	2
15-16	4	2	2
17-18	4	2	0
19-60	2	2	2
Mean no. documents:	6	5	4
Median no. documents:	3	3	2

<sup>a</sup>Based on search topics with overlap retrieval.

Table 6. Average recall and precision for overlap sets from the pilot and field studies

	Recall					Precision				
	Mean	SD	Min	Max	Med	Mean	SD	Min	Max	Med
<i>Field study</i>										
All relevant	0.13	0.16	0	0.83	0.08	0.81	0.31	0	1.00	1.00
Definitely relevant	0.17	0.19	0	1.00	0.10	0.69	0.34	0	1.00	0.80
<i>Pilot study</i>										
All relevant	0.12	0.19	0	1.00	0.06	0.92	0.18	0.33	1.00	1.00
Definitely relevant	0.16	0.20	0	1.00	0.10	0.75	0.36	0	1.00	1.00

There were 33% queries with five or more overlap papers, of which half had 100% precision and two-thirds had 90% precision. Even though few common items were found by terms and by citations, it was impressive that most of the overlap items were judged to be partially relevant or relevant.

#### Relevance odds and overlap retrievals

To compute the relevance odds of items derived from either one of the two search methods, as opposed to simultaneously from both search methods, the cross-product ratios were computed (Saracevic *et al.*, 1988a). To zero in on the different aspects of relevance, Saracevic and Kantor adopted three criteria of relevance. "Normal relevance" was computed as  $(R + pR)/(NR)$ , "strong relevance" was computed as  $(R)/(NR)$ , and "weak relevance" was computed as  $(R)/(pR + NR)$ . Specifically, "strong relevance" imposes a more rigorous requirement in that only those items of definite relevance can be used. Table 7a contains the data from which computations were based. Table 7b shows the relevance odds of items retrievable by one of the two search methods, as compared to those retrievable from both. Using these criteria of relevance for both field and pilot studies, it can be seen that the results of all six computations were statistically significant by a wide margin. In other words, in the open literature, if an item was derived from both types of search methods from MEDLINE and SCISEARCH, its relevance odds were improved by at least 5.7 times. Specifically, for a given query, the odds that an item was relevant or partially relevant (as opposed to not relevant) were increased by a factor of 5.7 for those retrieved items common to the outputs from both search outputs. For "weak relevance," the odds were nearly six times greater. Most astonishingly, for "strong relevance" the odds for items from both search methods were 8.4 times higher than for those retrieved by either method

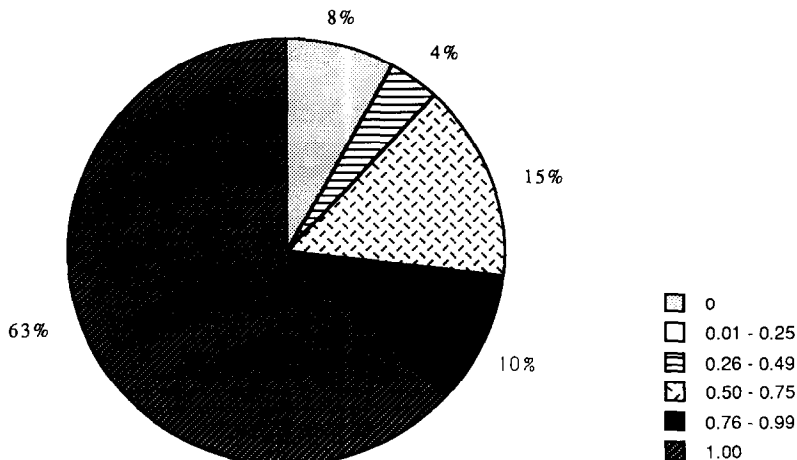


Fig. 3. Distribution of precision values of overlap sets.

Table 7a. Data for the computation of relevance odds of items from only one search method versus from both methods

	R	pR	NR	R + pR	pR + NR
<i>Field study</i>					
One method	1621	1299	2638	2920	3937
Both methods	197	43	38	240	81
<i>Pilot study</i>					
One method	524	311	549	835	860
Both methods	36	12	10	48	22

alone. Although the Saracevic/Kantor studies examined multiple retrievals in terms of searchers and not search methods, the relative order of magnitude with respect to the three relevance criteria in both studies was similar.

Tables 7a and 7b also show the log cross-product analysis from the pilot study data. The relevance odds for common items were increased by at least 2.69 times, which is much less than those found in the field study. The findings are consistent with the earlier comments that citation links in the small file were severely curtailed, thus adversely affecting the potential for overlap retrieval. Yet the overlap items still showed improved odds of being relevant when compared with items derived from either term or citation searching.

Finally, if only MEDLINE were searched, would there be advantages in consulting SCISEARCH on a limited basis to identify the duplicates? *What were the odds that a retrieved item from MEDLINE would be relevant or partially relevant (as opposed to not relevant) if it were only found in MEDLINE versus if it were also found in the citation database?* This question is different from the one posed in the previous section and in the section entitled "Relevance odds and search modes," where the comparison was made between overlap items and items from *either* database. For the following analysis, the focus is on the relevance odds of items retrievable from MEDLINE as a database. In other words, a comparison is made between items only found in MEDLINE and those also found in SCISEARCH, while other citation retrieval would not be considered. Table 8 presents the different relevance odds associated with three relevance criteria. The odds for the duplicates, as compared with those from MEDLINE alone, were 4.6 (6.32/1.36) times better, if normal relevance is considered. The performance of items retrieved only from SCISEARCH was considerably poorer. This indicates that given the choice of using MEDLINE alone versus either SCISEARCH or the pooled resources from both databases, there is indeed a higher probability of finding relevant items from MEDLINE. Most importantly, the odds that duplicates be definitely relevant were 6.4 (5.18/0.81) times higher.

Table 7b. Relevance odds of items from only one search method versus from both search methods

	Odds ratios	Log odds	SE	t value	Stat. sign
<i>Field study</i>					
Normal	5.71	1.74	0.12	14.12	Yes
Strong	8.44	2.13	0.14	15.62	Yes
Weak	5.91	1.78	0.13	13.38	Yes
<i>Pilot study</i>					
Normal	3.16	1.15	0.28	4.18	Yes
Strong	3.77	1.33	0.30	4.41	Yes
Weak	2.69	0.99	0.28	3.59	Yes

Normal relevance computed as:  $(R + pR)/NR$ .

Strong relevance computed as:  $R/NR$ .

Weak relevance computed as:  $R/(pR + NR)$ .

Table 8. Relevance odds for items retrieved from MEDLINE®, SCISEARCH®, and overlap

	MEDLINE® only	SCISEARCH® only	Overlap
Normal relevance	1.36	1.03	6.32
Strong relevance	0.81	0.61	5.18
Weak relevance	0.52	0.43	2.43

### *Practical considerations*

This section will discuss the practical implications of the combined use of term and citation searching conducted on two separate databases in operational settings. Several features are noted. MEDLINE searches are well-known to searchers and patrons alike. Since the database has been available for a long time, it is routinely used. SCISEARCH, however, is not a frequently used database in the health sciences libraries and it is an expensive database. As of May 1991, the hourly rate of usage is \$159.00 per connect hour. For those libraries with the hard copy of the *Science Citation Index*, the hourly charge is reduced to \$63.00. Even at the reduced rate, it amounts to \$1.00 per min, which is considerably higher than most MEDLINE searches.

Since citation searching is not often performed, searchers may be less familiar with search features unique to SCISEARCH. For example, for a cited reference, one must have the correct bibliographic citation in order to formulate the proper search string. To do so, one should conduct an online verification of the exact bibliographic data of the seed document supplied by the patron. If the searcher is unfamiliar with the proper placement of punctuation, format, and author initials used, one could easily misinterpret a zero hit as "no citing paper available," when an incorrect search string was the culprit. For example, "no hit" would result from a search statement such as "CR=WARREN KS 1980, V303, P203, NEW ENGL J MED," whereas at least 36 citing references would be retrieved by "CR=WARREN KS, 1980, V303, P203, NEW ENGL J MED." Additionally, the abbreviations used for a journal title are not unified, neither are names for authors. Thus, the author's name should be "expanded" to obtain all the variant forms used.

On the other hand, many end-users are unfamiliar with SCISEARCH. Occasionally, there is mistrust of a strange database, as was shown by a patron's comments. Indication of misunderstanding on the part of the patron was also noted when searches for the field study were collected. One patron dismissed the entire citation search out of hand because the printout format was different from the familiar MEDLINE output. The amount of data included in SCISEARCH and the order of presentation are quite different from that of MEDLINE.

Initially, it was feared that patrons might not supply the bibliographic citation of one or two appropriate papers on the search topic. Relevant retrieval from citation searching depends solely on good seed documents. Without them, citation searching simply cannot perform well. Fortunately, in this project, most participating patrons were able to supply appropriate seed documents. A few even gave two, and two users offered three. This fact is probably due to the cooperation of the participating librarians. In general, if patrons have enjoyed quality service in their libraries, there is a stronger tendency for the patrons to cooperate.

Although the number of overlap items from the two databases is generally small, duplicate records do appear in the printouts. A prototype program has been produced by the project to merge two sets of records. It has the options to remove the duplicates, and to identify the overlap items. Thus, a fairly simple front-end program could be built into existing search programs. However, for this study, the librarians would have to postprocess the output by running the program. It involves an added step in their normal search procedure. Therefore, the program was not utilized. As of this writing, DIALOG® has introduced the duplicate records removal feature on many databases. By sorting the merged list, one could identify the common items. Unfortunately, for the biomedical literature user,

this convenience is limited only to the use of MEDLINE via DIALOG. Since many health science libraries access MEDLINE via NLM or BRS®, this feature is of no help. Therefore, the combined use of MEDLINE and SCISEARCH is possible, but the mechanism needs to be refined and made easier for the online searcher.

### *Topical characteristics*

To gain insight into the types of queries which could benefit most from the option of citation searching, the content of the queries was examined in light of levels of recall and precision. Three levels of retrieval performance were arbitrarily determined. Values above 75% were considered "high", 40% up to 75% were "medium", and below 40% were "low." Based on the levels of recall and precision performance from citation retrieval, each search topic was classified into one of the nine cells. As a result, the queries were distributed among nine categories. For example, one cell contained queries which achieved high recall and precision from SCISEARCH. Each group of topics was examined for any consistent characteristics. Twenty-six percent of the queries belonged to the largest group with relatively high recall and precision. Although there was a variety of topics, they appeared mostly to be well-defined in scope. A sizable number related to questions concerning treatment and therapy. SCISEARCH appeared to be least effective in queries that were very specific and tended to lie within a narrow scope. As expected, those queries impinging on the behavioral sciences appeared to derive the most benefit from citation searching (Goffman & Pao, 1980; Kochtanek, 1982). This situation could have been reexamined if *Social Science Citation Index* was also searched.

Topics associated with different levels of precision were much more difficult to generalize since some queries retrieved a large number of postings. Yet, even if most retrieved items were relevant, higher precision is very much dependent on individual needs. In one case, one user commented that although many items in the printout were useful, he was only interested in two or three items to use in preparing for a talk. The attempt to identify query characteristics was not entirely successful. In future studies, more creative methods should be used to identify and to distinguish differences in search topics in order to test the hypothesis that topical characteristics are related to levels of retrieval performance associated with the two search methods.

## DISCUSSION

It was noted earlier that only a few published studies directly compared the results of term and citation retrieval for the same set of queries. Operationally, these two methods of searching are done by two search modes. The underlying assumption is that these two search methods are based on two different criteria of retrieval, which in turn represent two distinct types of subject relevance (McCain, 1989; Pao, 1986).

Several classes of variables are known to affect retrieval effectiveness. Obviously, controlling of all such variables is difficult, if not impossible, in the field setting. This study was designed to generate data to compare with data produced in the pilot study, and emphasis was placed on determining the appropriate magnitude of incremental contribution by each retrieval method with real-life library situations. As a result, a wide variety of biomedical topics which were avoided in the pilot study was used. Topics were only limited by the requirements articulated by the user. Since most were higher precision searches, the control that could be introduced by the requirement of high recall for all searches, such as those used by McCain, was sacrificed. Unlike the use of a file on a narrow topic in the earlier study, MEDLINE with its broad subject coverage was used. Contrasting with the use of a single searcher in the pilot study, the 89 topics were searched by several searchers. Therefore, in all likelihood, the recall from either term or citation retrieval was less than what could have been retrieved by either search method.

On the other hand, intervention brought on by the study was kept to a minimum. Searches conducted at the participating libraries were used. No attempt was made to influence the searchers in their MEDLINE searches in terms of cost or time. Searchers were not instructed to aim for total recall from MEDLINE, and users were left to identify as many

key contributions for citation search keys as they wished. Specific to citation searching, it was noted earlier in the pilot study that the seed document used could dramatically alter the search result. Yet, in this field study, the appropriateness of the seed document used was not assessed. An assumption was made that given a reasonable sample size obtained in near-naturalistic conditions, the search results would be likely to resemble those typically found in other health sciences online searching environments.

Furthermore, evaluation of the relevance of items retrieved was performed by the requestors themselves instead of surrogates. Although the consistency of a single evaluator was sacrificed, the study again gained accuracy from the near-realistic condition of information encounter faced by real users with real information needs. A total of 7208 separate relevance judgments were made. However, there was still an element of self-selection of users, since they had to participate in the evaluation process.

### *Contributions from citation searching*

Overall confirmation of the pilot study was found. No surprisingly, MEDLINE contains a large proportion of relevant materials for biomedical topics. Yet this study confirms earlier findings that most online searches in the health sciences libraries could benefit from even one additional citation search. From the sample of 89 topics collected from real users, an average of 24% relevant materials not retrieved from MEDLINE was found. Citation retrieval from the subject-specific file of enteral hyperalimentation added only 14% relevant items. Although in both studies, the majority of citation searches was restricted to one seed reference, citation searching in a multidisciplinary database produced significantly more relevant items.

McCain (1989), however, reported a 33% incremental relevant retrieval from citation searching. She reasoned that effective citation retrieval requires that: (a) At least a few key contributions to the research topic exist; (b) the importance of these papers is generally recognized by researchers; (c) citation of these key papers is an expected practice of that discipline; and (d) sufficient time has passed after the publication of these key papers in order to accumulate a body of citations. Therefore, the number of documents retrieved, and the number of relevant retrievals can vary widely depending on the number and the appropriateness of the seed documents chosen. While no attempt was made to assess objectively the appropriateness of search keys in these studies, there was a marked difference in the number of seed documents used. In comparing the performance of descriptor and citation searching, McCain attempted high recall from both types of retrieval and used three descriptor-based files and two citation databases. She was able to engage her users to produce two to eight key documents for each topic, and she used four or more in seven of the nine topics. In the present study, the focus was on searches originating from real needs and being made to meet real requests. A search sample of 89 topics was used. In the course of the study, every effort was made to minimize intrusiveness to both the searcher and the user. Attempts were made to capture term and citation searching found in real-life library situations. Most often, requests in the health sciences libraries were found to be high-precision searches. When asked, users tended to give a single seed document. Therefore, the differential of citation retrieval as compared with those reported by McCain may be partially explained by the difference in the number of seed documents used. Nevertheless, the relationship between the number of seed documents and the amount of unique contributions from citation searching should be addressed in future studies.

### *Precision performance*

While precision from either MEDLINE or from SCISEARCH was lower than that found in the pilot study in absolute number of retrieved items, fewer than 16 new items were added by citation retrieval to two-thirds of the search topics, half of which were relevant. Practically speaking, the onus of scanning additional items from SCISEARCH would be relatively minor. Thus, in the health sciences arena, while at the moment at least two separate databases must be searched, an added citation search can be expected to have definite payoff. Unfortunately, due to the added cost constraint, resistance in using an ad-



ditional citation search was indicated. The increasing availability of CD-ROM versions of these databases might encourage more usage.

### *Relevance odds of overlap sets*

The most important finding from this study was that the odds that overlap items would be partially relevant or relevant, as opposed to not relevant, were improved by six times, as compared with items retrieved by using only either one of the search methods alone. The relevance odds increased to 8.4 times if items of definite relevance were needed. In the past, only broad agreement existed that overlap items from multiple retrievals tended to be relevant. The strong evidence found is also supported by analysis of the search data from the pilot study. Even higher precision requirements could be satisfied by such duplicates. Practical implications in terms of interface design of search softwares are obvious. As the size of databases continues to grow, high precision retrieval of definite relevant items assumes more importance. If present search systems are able to eliminate duplicate retrieved items, it should be a relatively simple task to extend this feature to tag such duplicates.

### *Topical orientation*

Obviously, citation retrieval did well for interdisciplinary and emerging topics. Topics of well-defined scope in treatment and drug therapy tended to derive higher recall and precision from citation searching. Unfortunately, the results from examining the relation between topical characteristics and higher payoff from an additional citation search were inconclusive.

### *Log cross-product analysis*

Methodologically, the cross-product analysis produced more meaningful and more precise results. In the re-analysis of data from the pilot study, this analytic technique confirmed that in-depth subject indexing of a small file indeed improved term-based retrieval substantially. More importantly, the intuitive notion that MEDLINE contains most of the relevant items for biomedical topics as compared with SCISEARCH was supported by improved odds of a specific dimension. Furthermore, instead of another general observation that the overlap items from both databases tended to produce more items of definite relevance, the analysis showed that the odds were improved by an astounding 8.4 times. A more precise comparison is made available as a result of the micro-level analysis based on the relevance odds of each item retrieved.

## CONCLUSION

Enough has been written to suggest that different criteria of relevance may be at work when different access points are used or when different searchers search the identical topic. A recent paper reporting on MEDLINE and full text searching also bolsters the complementary nature of using different subject access (McKinin, Sievert, & Mitchell, 1991). The exact nature of the relationship between the document content and the different terms used to represent them remains unclear. Neither are the underlying principles of citation linkages of documents illuminated. Based on a single high recall subject-based search, Harter offered a tentative explanation on why many searches reported low recall (Harter, 1990). He explained that there are many reasonable candidate search elements and search terms for any given topic. A normal search strategy would usually use only some of the terms, logically producing lower recall. The degree of redundancy introduced from different terms is unknown. Clearly the complex interplay of language and meaning has significant impact on retrieval. These findings underlie the need for additional experimentation and further research.

*Acknowledgements*—The author gratefully acknowledges the help from the staff of the Taubman Medical Library of the University of Michigan, the Medical Library of the Cleveland Clinic Foundation, the Medical Library of the McLaren Hospital, and the Medical Library of Rush University. Without their cooperation, this study would not have been possible. Dr. Judith Weedman of UC Berkeley has provided invaluable moral and material help in the writing of this paper. Helpful comments from the reviewers are also acknowledged.

## REFERENCES

- Bates, M.J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science*, 37(6), 357-376.
- Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-424.
- Bichteler, J., & Eaton, E.A. (1980). A combined use of bibliographic coupling and co-citation of document retrieval. *Journal of the American Society for Information Science*, 31(4), 278-282.
- Blair, D.C., & Maron, M.E. (1990). Full-text information retrieval: Further analysis and clarification. *Information Processing and Management*, 26, 437-447.
- Borgman, C.L. (1986). Why are online catalogs hard to use: Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37(6), 387-400.
- Chapman, J., & Subramanyam, K. (1981). Co-citation search strategy. *Proceedings of the 2nd National Online Meeting*, Medford, NJ, Learned Information (pp. 97-102).
- Goffman, W., & Pao, M.L. (1980). Retrieval of biomedical information for emerging interdisciplinary problems. *Proceedings of the 4th International Congress on Medical Librarianship*, Belgrave, Nova Gradiška (pp. 39-50).
- Harter, S.P. (1990). Search term combinations and retrieval overlap: A proposed methodology and case study. *Journal of the American Society for Information Science*, 41(2), 132-146.
- Katzer, J., McGill, M.J., Tessier, J.A., Frakes, W., & DasGupta, P. (1982). A study of the overlap among document representation. *Information Technology: Research and Development*, 2, 261-274.
- Kochtanek, T.R. (1982). Bibliographic compilation using reference and citation links. *Information Processing and Management*, 18(1), 33-39.
- McCain, K.W. (1989). Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution. *Journal of the American Society for Information Science*, 40(2), 110-114.
- McCain, K.W., White, H.D., & Griffith, B.C. (1987). Comparing retrieval performance in online data-bases. *Information Processing and Management*, 23(6), 539-553.
- McKinin, E.J., Sievert, M.E., & Mitchell, J.A. (1991). The MEDLINE/Full Text Research Project. *Journal of the American Society for Information Science*, 42(4), 297-307.
- Pao, M.L. (1986). Comparing retrievals by keywords and by citations. *Proceedings of the 7th National Online Meeting*, Medford, NJ, Learned Information (pp. 341-346).
- Pao, M.L., & Fu, T.W. (1985). Title retrieved from MEDLINE and from citation relations. *Proceedings of American Society for Information Science*, 22, 120-125.
- Pao, M.L., & Worthen, D.B. (1989). Retrieval effectiveness by semantic and pragmatic relevance. *Journal of the American Society for Information Science*, 40(4), 226-235.
- Salton, G. (1971). Automatic indexing using bibliographic citations. *Journal of Documentation*, 27, 98-100.
- Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving: III. Searchers, searches, and overlaps. *Journal of the American Society for Information Science*, 39(3), 197-216.
- Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving: II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3), 177-196.
- Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. (1988). A study of information seeking and retrieving: I. Background, and methodology. *Journal of the American Society for Information Science*, 39(3), 161-176.
- Shaw, W.M. (1990). Subject indexing and citation indexing, II. An evaluation and comparison. *Information Processing and Management*, 26(6), 705-718.
- Sparck Jones, K. (1988). Fashionable trends and feasible strategies in information management. *Information Processing and Management*, 24(6), 703-711.
- Sparck Jones, K. (Ed.). (1981). *Information retrieval experiments*. London: Butterworths.
- Sparck Jones, K., & Van Rijsbergen, C.J. (1976). Information retrieval and test collections. *Journal of Documentation*, 32, 59-75.
- White, H.D., Griffith, B.C., Cowen, J.A., Selinger, N.E., & Steere, D.T. (1984; January). *Evaluation of the National Library of Medicine's programs in the medical behavioral sciences. Quality of indexing: The development and testing of a behavioral science literature* (Report to the NLM). Philadelphia: Drexel University.
- Worthen, D.B. (1979). Characteristics of enteral hyperalimentation literature. *Journal of Parental and Enteral Nutrition*, 3(6) 465-467.