

---

# Separate but equal? a system comparison study of MEDLINE's controlled vocabulary MeSH

By Karin Schreier Hallett, M.A., M.S.  
User Services Librarian

Andrew L. Todd Library  
P.O. Box 13  
Middle Tennessee State University  
Murfreesboro, Tennessee 37132

---

This study tested the effect of controlled vocabulary search feature implementation on two online systems. Specifically, the study examined retrieval rates using four unique controlled vocabulary search features (explode, major descriptor, descriptor, subheadings). Each search feature was applied to nine search queries obtained from a medical reference librarian. The same queries were searched in the complete MEDLINE file on the Dialog and Ovid systems. The unique records, i.e., those records retrieved in only one of the two systems, were identified and analyzed. Dialog produced equal or more records than Ovid in nearly 20% of the queries. The study demonstrated that users need to be aware of system-specific designs that may require differing input strategies across different systems for the same unique controlled vocabulary search features. The paper concludes by making recommendations and suggestions for future research.

---

## INTRODUCTION

Librarians may choose from several vendors when purchasing access to the MEDLINE database. Two of the more popular systems are Ovid (Ovid Technologies, Inc.) and Dialog (The Dialog Corporation). An increased dependence on accessing large amounts of information electronically, in a context of diminishing resources, makes evaluation of the degree to which systems provide equal access to data a vital enterprise. This study has examined the extent to which Dialog (File 155) and Ovid (File Mesz) provide access to the MEDLINE database via the Medical Subject Headings (MeSH). Both files are directly available from Dialog and Ovid, respectively. The study sought to answer two research questions: (1) What, if any, are the general differences between controlled vocabulary system implementations on Ovid and Dialog? and (2) What, if any, are the impacts of each of the differing controlled vocabulary search features upon retrieval rates? If there are indeed differences in retrieval rates due to the implementation of unique controlled vocabulary search features—such as explode, major descriptors, descriptors, and subheadings—these search features become important variables in the systems evaluation process.

## RELATED LITERATURE

While the function of controlled vocabulary is to facilitate the retrieval of relevant information, electronic

information retrieval systems have compounded the amount of information available. At the same time, studies on the utility of controlled vocabulary in electronic environments have proliferated. The literature relevant to this study falls into two broad areas: (1) general descriptions of MEDLINE implementations on various systems and (2) system-specific search features of MeSH on Ovid and Dialog. Comparisons of MEDLINE implementations on different systems, CD-ROM or online, have focused on user utility of search features. No published studies have compared retrieval rates for MEDLINE on Dialog and Ovid.

## GENERAL DESCRIPTIONS OF MEDLINE IMPLEMENTATIONS ON VARIOUS SYSTEMS

MeSH is perhaps the most popular controlled vocabulary system currently operating in an online environment. It is used on, among other databases, MEDLINE. General evaluations and comparative studies of implementations of the MEDLINE file on different systems have focused on the special search features available through MeSH. While MEDLINE has been implemented on many different systems, its advantages and disadvantages across systems remain unclear. Users often have individual search strategy preferences depending on their skills, training, and familiarity with the system [1, 2]. Although there are multiple variables

to consider when determining the preferred implementation of MEDLINE, user skill has been shown to be the most important; MEDLINE implementation should thus be determined by user needs [3]. Moreover, user searching skills and the ability to take advantage of MEDLINE's unique search features is critical to performance outcomes [4].

## MeSH SEARCH FEATURES ON OVID AND DIALOG

Subheading searching and explosions have been identified as the most powerful controlled vocabulary search capabilities on Dialog and Ovid. A 1986 study compared searching MEDLINE subheadings on BRS, Dialog, and MEDLARS and listed tips on searching subheadings [5, 6]. In the same year, another study compared the search capabilities of the Dialog Medical Connection with those of other systems, including MedBase, Grateful Med, BRS/Colleague, and Paper-Chase, for user-friendliness. The authors concluded that while the system's commands were easy to learn and use, its menu mode should be enhanced by including more guidance for the use of subheadings and explode capabilities of MeSH and they recommended adding search levels to the Medical Connection menu mode to allow the user to browse both titles and descriptors and to provide guidance to subheadings and explodable terms [7].

A later study compared BRS, Data-Star, and Dialog in relation to pricing, updates and SDIs, record format, name searching and bibliographic verification, free-text searching, limiting, the online thesauri as well as general system features, and examined descriptor and subheading searching. The study found that overall, depending on the search strategy, one system might perform better than the other [8].

## METHOD

To determine system performance quality, this study examined possible impacts on retrievability due to the unique, system-defined controlled vocabulary search features of the online versions of Ovid and Dialog. The complete MEDLINE file was searched on March 22, 1997, using each system's command language interface. Nine search questions submitted by users of the Preston Medical Library at the University of Tennessee Medical Center in January 1997 were obtained from a librarian. All search strategies were composed prior to going online. The search strategies were composed of descriptors and subheadings taken from the MeSH indexing language. While this restriction to controlled vocabulary terms may have produced an unrealistic search situation because users typically search with a combination of natural language and controlled vocabulary, as did the Preston Medical librarian, it

**Table 1**  
Search topics

1. Attorney wants to find articles related to mental disorders/memory/cognitive disorders related to astrocytomas. Someone is contesting last will and testament.
2. Resident wants articles related to neurofibromatosis and "any" type aneurysm.
3. Psychiatrist requests information on use of chromium or chromium picolinate in weight control/weight loss.
4. Emergency room physician wants information on patient satisfaction with ED/ER and processing of patient through the ED/ER, i.e., long waits etc.
5. Resident wants articles related to patient with heart valve prosthesis undergoing non-cardiac surgery and anticoagulant therapy—post-op.
6. Physician wants references to diencephalic or hypothalamic syndromes.
7. Physician wants articles on breast fibroadenoma, "atypical" fibroadenoma, or carcinoma fibroadenoma.
8. Physician wants information on the use of chelation therapy to treat atherosclerosis.
9. Physician wants information on melatonin and its effect on sleep/sleep disorders.

served to highlight the effects of controlled vocabulary implementation on system performance. An effort was made, however, to preserve the originality of the queries as formulated by the librarian, i.e., the search terms used by the librarian were applied in this study's strategies provided they were MeSH terms. Further, while the librarian restricted all her searches to the last five years of the MEDLINE file (1992 to January 1997), the search strategies composed for this study added MeSH terms in usage by the National Library of Medicine (NLM) prior to those dates to search the complete MEDLINE file. These, as well as the subheadings, were taken from the printed MeSH. The search topics are listed in Table 1.

To retrieve a workable sample of documents, defined as 100 or fewer documents, searches were limited to the English language and the human check tag. Further, to eliminate a possible problem of a difference in loading dates between the systems, all searches were limited to the publication years 1966 to 1990. Despite these limitations, however, some of the searches retrieved more than 100 documents. In those instances, searches were further limited by publication year. Care was taken to apply the same search strategy on each system to allow a comparison of results.

To each of the nine queries, the following unique controlled vocabulary search features were applied for a total of thirty-six queries per system: *Explode*, searched all narrower terms of a subject heading in the MeSH hierarchy from up to twelve levels; *Major Descriptor*, searched MeSH terms preceded by an asterisk and representing the core concept of a document; *Descriptor*, searched all MeSH terms; *Subheadings*, searched for very specific aspects of a subject. For comparison purposes of the search results, the retrieved records were sorted by author and the accession number, author, and title of each record printed.

The unique records from each search, defined as those not retrieved in one or the other system, were

identified and recorded. To calculate the retrieval failure rate for each controlled vocabulary search feature of each query, the number of unique records retrieved was divided by the highest number of records retrieved. The resulting number was then multiplied by 100 to express the retrieval failure rate in percentage. To verify that the unique records were not indeed on the other system, a known-item search was conducted. The accession number, author, title, and descriptor fields of all unique records were then printed to determine categories for reasons of failure. The categories were determined by the following procedure:

- The descriptor field of all unique records was closely examined to see if and how the search terms were listed. If a search term was not listed in the descriptor field, the search term's narrower and broader terms were identified online and checked against the descriptor field of the unique record. If a unique record was retrieved with a search term's narrower or broader term, both systems were further tested employing the same controlled vocabulary search feature to identify if either system required a different search strategy than was initially entered.
- If the search terms were listed and did not indicate any reasons for failure, a set of elimination strategies was executed online. These strategies searched the pertinent queries again by checking after each step whether the unique document was part of the newly retrieved set. For example, the original search query was combined with the accession number of the unique document. Then the original search query was limited to the human check tag and this newly created set was combined with the accession number of the unique document. The original set was then limited to English language documents and this set as well was combined with the accession number of the unique document, and so forth. If the elimination strategies did not answer the question of retrieval failure, the technical support representatives of either Dialog or Ovid were contacted for further assistance.

## FINDINGS

The 36 searches conducted on each system retrieved a combined total of 2,204 documents. Of these, 116 documents (5.26%) were unique, or retrieved in only one of the two systems (Dialog). For 7 of the 36 searches (19.4%), Dialog retrieved more documents. Interestingly, most of the unique document retrievals occurred when limiting searches to major descriptors. Moreover, the research also revealed that for each controlled vocabulary search feature applied to Query 9, there was a discrepancy in the retrieval rates between the systems. The other discrepancies occurred with Query 2 (major descriptors) and Query 6 (major descriptors and descriptors). A known-item search of these unique

**Table 2**  
Distribution of unique documents by search feature

CV search feature	Search query	Number of unique documents (percentage)
exp	9	1 (0.86%)
maj	2, 6, 9	112 (96.55%)
de	6, 9	2 (1.72%)
sub	9	1 (0.86%)

documents showed that, while not retrieved in the queries posed, all were available on Ovid. Table 2 is a distribution of the number of unique documents by controlled vocabulary search feature.

While retrieval rates differed for seven of the search queries, by far the majority of unique documents were retrieved when the major descriptor controlled vocabulary search feature was applied. The reason for differing results with the application of this search feature could be attributed to Dialog's policy of double-posting MeSH terms, also known as combination parsing, a combination of both word and phrase parsing. A multi-word descriptor phrase, for example, would be indexed by phrase as well as by word (with the exception of stop words). For example, Search Query 2 limited to major descriptor was entered on Dialog as *fibroadenoma/maj* and (*carcinoma/maj* or *breast diseases/maj*), retrieving not only documents with carcinoma as a major descriptor but also documents listing carcinoma as part of a major descriptor phrase like *carcinoma, infiltrating duct*. In order to avoid retrieving single-word MeSH terms as part of a bound phrase in Dialog, a searcher would have to limit the search term to full descriptor, i.e., enter it as *carcinoma/df,maj*. All three cases in which the search limited to major descriptors retrieved more documents in Dialog than Ovid were re-entered applying the full descriptor limit. The results in both systems were now the same.

Thus, different system policies on indexing necessitate different search strategies for a system. While in Ovid the descriptor field is phrase indexed only, in Dialog it is both phrase and word indexed. To compensate for the differences in search results, a searcher would either have to limit a single-word descriptor search to full descriptor in Dialog, or to broaden a single-word descriptor search by applying the heading word post qualifier (.hw) in Ovid. In Ovid, this would be a two-step process as the system does not provide the possibility to directly apply the post qualifier and limit the search to major descriptor at the same time.

While more documents were retrieved in Dialog due to its policy of double-posting descriptor terms, this higher retrieval rate did not mean that the extra documents were relevant. On the contrary, in all likelihood they were irrelevant because not only documents with the single-term descriptor but also documents with that search term as part of a descriptor phrase

were retrieved. Double-posting will always automatically broaden a search. A searcher using MeSH terms might not find this advantageous. MeSH in itself is a highly developed indexing language, which will facilitate precise search results if consulted before searching.

Regarding the reasons for retrieval failure with the application of the remaining three controlled vocabulary search features, neither an examination of the descriptor fields of the unique records nor the online elimination strategies revealed any reasons for failure to retrieve. The elimination strategies did clarify, however, that neither limiting to human check tag nor English language documents caused non-retrieval; rather, only after imposing the publication year limit were the records no longer retrievable. There were four unique document occurrences; one of the documents was retrieved in three of the queries with a second document retrieved in one query. A technical error was assumed and an Ovid technical support representative was contacted for further assistance. The Ovid representative explained that the occurrences of unique records were due to the fact that those records were of chapters in monographs. The NLM has not indexed monographic chapters under publication year field, instead the library listed the publication year in the source field and the NLM call number field. Since, for the unique records, the publication year field was left blank, Ovid did not retrieve them after limiting to publication year. Dialog, on the other hand, treated these records like journal articles and pulled their publication dates from the NLM call number field. While the source field was searchable in Ovid, searching it for a range of years (e.g., 1966-1990.so) was impossible. Rather, in order to retrieve the records of monographic chapters for this time period in Ovid, a searcher would have to enter a search string ORed for each publication year (e.g., 1966.so OR 1967.so OR 1968.so OR 1969.so and so forth).

While this system difference occurs when searching with unique controlled vocabulary search features, it is important to note that differing retrieval results would also occur when performing, for example, an unqualified search. Hence, this category for reason of failure is by no means attributable to controlled vocabulary searching alone. Instead, in examining the possibilities for differing retrieval rates, the following conclusions can be drawn:

- Information professionals cannot assume symmetrical retrievability from one database on different systems when applying the same controlled vocabulary search features. It is important to test this assumption. As this study demonstrates, implementation of controlled vocabulary can influence search results.
- Word and phrase parsing, or double-posting, is an important variable when searching on controlled terms. While systems employ the same controlled vo-

cabulary search features, different results may be retrieved by different systems searching the same database. The searcher needs to be aware not only of the system-specific search features, but also of system policies on how data is made available. For example, a searcher can compensate for the differences in parsing by altering the search strategies.

## CONCLUSION

Over the last thirty years, information retrieval systems have become more sophisticated, more powerful, more user friendly, and certainly more prevalent. The research on controlled vocabularies, and specifically comparison evaluation studies on implementations of MEDLINE across different systems, have been ongoing. Central has always been the question: What system is "best" relative to the research situation and the user? This question is now more relevant than ever considering the expanding diversity of user populations and research settings. How can system designers best make their information available to a broad spectrum of users? Discussions of index languages and their utility across systems can be more important than the information itself.

What are the implications of this study? Searchers typically assume that they should be able to retrieve the same information across different systems. They rely on the assumption that because the indexing of the documents is undertaken by the NLM, either system will retrieve the same records if the same query is posed applying the same unique controlled vocabulary search features. As this study has shown, this is largely the case. Retrievability, however, is not just a question of numbers of documents. From the user's perspective, it is a question of relevant documents retrieved. For example, while the 36 searches conducted in this study yield 2,204 documents, 116 of those documents are unique on Dialog. Yet, in all likelihood, most of these unique documents, due to the fact that they are retrieved as a result of Dialog's policy of double-posting, are not relevant.

This study is another in a series of studies testing the effectiveness of controlled vocabulary searching in online retrieval systems. Asymmetrical search results are retrieved due to the searcher's lack of familiarity with how controlled vocabulary is implemented on the two systems. Searchers need to be aware of and acquire a working knowledge of how one file is made searchable on different systems. Particularly in the medical profession, obtaining all pertinent documents is crucial.

Ovid should change implementation of the MEDLINE file by making book chapters searchable as articles; (1) to eliminate the danger of a searcher unknowingly not retrieving a document, and (2) to prevent a searcher aware of the file implementation dif-

ference from having to add an extra search step in order to include book chapters when limiting to publication year. By the same token, Dialog should change implementation of the MEDLINE file to eliminate double-posting of descriptor terms. While double-posting increases the number of search results, it does little to improve retrieval of relevant documents.

While this study was limited to only two systems and therefore conducted on a rather small scale, it did show that by no means do two systems (while offering the same unique controlled vocabulary search features) retrieve precisely equal numbers of records. Future studies might compare multiple systems, perhaps even across different interfaces.

#### ACKNOWLEDGMENT

I am indebted to Carol Tenopir, P.h.D., Peiling Wang, P.h.D., and David Penniman, P.h.D., all of the University of Tennessee's School of Information Sciences, for their comments and suggestions throughout the project. I would also like to thank Doris Prichard and Connie Littleton, librarians at the Preston Medical Library at the University of Tennessee, for sending me

several "real-life" questions, of which I chose nine to test in this study.

#### REFERENCES

1. BAKKER S, BLEEKER A, VAN DER BURG J, DIJKMAN J ET AL. MEDLINE on CD-ROM: a comparison. *Online Rev* 1989;13(1):39-50.
2. BONHAM MD, NELSON LL. An evaluation of four end-user systems for searching MEDLINE. *Bull Med Libr Assoc* 1988 Jan;76(1):22-31.
3. DELOZIER EP, LINGLE VA. MEDLINE and MeSH: challenges for end users. *Med Ref Serv Q* 1992 Fall;11(3):29-46.
4. MCKIBBON KA, HAYNES RB, WALKER DILKS CJ, RAMSDEN ME, ET AL. How good are clinical MEDLINE searches? a comparative study of clinical end-user and librarian searches. *Comput Biomed Res* 1990 Dec;23(6):583-93.
5. SNOW B. MEDLINE subheadings: review and update. *Database* 1986 Aug;9(4):79-88.
6. SNOW B. MEDLINE subheadings revisited. *Database* 1986 Oct;9(5):63-71.
7. KWAN J, WATSON JL, DENNEY K. Dialog Medical Connection—an evaluation. *Online* 1987 Nov;11(6):32-8.
8. HUNTER J. MEDLINE on BRS, Data-Star and Dialog—does it make a difference? *Database* 1991 Feb;14(1):43-50.

*Received November 1997; accepted March 1998*