

---

# Creating local bibliographic databases: new tools for evidence-based health care\*

By Jeremiah H. Sable, M.D.

Beth G. Carlin, M.A.L.S.

James E. Andrews, M.L.S.

MaryEllen C. Sievert, Ph.D.

Department of Health Management and Informatics  
School of Medicine

University of Missouri–Columbia

324 Clark Hall

Columbia, Missouri 65211

---

The Internet has created new opportunities for librarians to present literature search results to clinicians. In order to take full advantage of these opportunities, libraries need to create locally maintained bibliographic databases. A simple method of creating a local bibliographic database and publishing it on the Web is described. The method uses off-the-shelf software and requires minimal programming. A hedge search strategy for outcome studies of clinical process interventions is created, and Ovid is used to search MEDLINE. The search results are saved and imported into EndNote libraries. The citations are modified, exported to a Microsoft Access database, and published on the Web. Clinicians can use a Web browser to search the database. The bibliographic database contains 13,803 MEDLINE citations of outcome studies. Most searches take between four and ten seconds and retrieve between ten and 100 citations. The entire cost of the software is under \$900. Locally maintained bibliographic databases can be created easily and inexpensively. They significantly extend the evidence-based health care services that libraries can offer to clinicians.

---

## INTRODUCTION

The Web has revolutionized the relationship between information seekers and information of every sort. Web interfaces to online databases, such as PubMed, make it possible to perform literature searches from almost any computer connected to the Internet. All of these searches are possible without any specialized training in information retrieval. However, while the Internet has increased everyone's access to information, effective literature searches still usually require the expertise of information professionals.

Various attempts have been made to simplify MEDLINE searching, such as PubMed's Clinical Queries, which have "research methodology filters" for therapy, diagnosis, etiology, and prognosis. The Hardin Li-

brary at the University of Iowa has incorporated these research methodology filters into a simple and elegant Web page front end to PubMed for searching the dermatology literature [1].

Many libraries use hedges, which combine text words and indexing terms to find the elusive literature [2]. Another familiar service provided by the National Library of Medicine (NLM), called selective dissemination of information (SDI), provides a means for searchers to update their hedged searches without the drudgery of entering the terms each time a search is run. This method provides a workable means of getting at the needed literature.

Over the last ten years, there have been changes in NLM's SDI services. However, there still is not an easy way to present the SDI results to clinicians. Furthermore, the monthly SDI updates to a hedge remain isolated from each other and do not form a single, searchable database.

---

\*This work was supported by the National Library of Medicine grant 5 T15 LM07089-05.

Of the many factors that have contributed to the Internet information explosion, two are particularly relevant to medical librarians: (1) the simplicity of using Web browsers and (2) the ease with which databases can be connected to Web browser front ends. These features of Web technology make it possible for medical librarians to create easy-to-use, locally maintained bibliographic databases to serve the specialized needs of clinicians. Librarians can use local databases to compile the results of monthly SDI searches and to present the search results to clinicians through a Web browser interface that requires little skill in literature searching.

A simple method for librarians to create Web-searchable bibliographic databases was developed. The methodology used inexpensive, readily available software and required minimal programming skills. As a demonstration of this methodology, the project staff created OutcomeLit, a bibliographic database of outcome studies.

## METHODS

The method of creating a locally maintained bibliographic database involved twelve relatively simple steps:

1. Develop a hedge search to retrieve relevant citations.
2. Search an online database to retrieve the citations.
3. Save the search results.
4. Import the search results into a reference management software application.
5. Reformat the citations as desired.
6. Export the formatted citations as tab-delimited text files.
7. Import the text files of citations into a database management software application.
8. Make the database Web-searchable by installing the ODBC database driver specific to that database application.
9. Create a Web page front end for users to search the database.
10. Write a few short scripts to connect the database to the Web page front end.
11. Use a Web server to publish the Web page front end on the Internet.
12. Update the database monthly to keep it current.

These general steps can be accomplished in numerous ways by using different software tools. For example, step 2 can utilize any bibliographic database search engine, such as Ovid, PubMed, or Internet Grateful Med. Step 4 can be accomplished with various reference management applications, such as EndNote, ProCite, or Reference Manager. Numerous database applications can be used for step 7, such as Oracle, Sybase, or Microsoft Access. Many different scripting languages are available for step 10, such as

JavaScript, Perl, or Microsoft Internet Database Connector (IDC).

Various combinations of these or other tools could be used successfully. The method described here used only software that was (1) non-customized and off-the-shelf, (2) inexpensive, and (3) easy to use. The following software and techniques were used.

### 1. Creating a hedge search strategy

The sample bibliographic database consisted of citations of outcome studies of clinical process changes and interventions. The hedge search strategy to retrieve these citations used three concepts: (1) an intervention concept, (2) a type-of-article concept, and (3) an outcome-management concept. For the intervention concept, terms such as "critical pathways," "case management," and "multidisciplinary teams" were used. The type-of-article concept was used to retrieve citations of studies and other articles that reported data. Some terms used for this concept were "retrospective study," "methodology," and "study design." To express the outcome management concept, terms such as "hospital costs," "benchmark," and "satisfaction" were used. For each concept, both text words and MeSH terms were used, a total of 128 in all. The three concepts were then combined using the Boolean AND to create the hedge.

### 2. Retrieving citations from a database

MEDLINE was searched from 1988 through 1998 using the Ovid Windows client software. The search results were limited to articles published in English and then limited by publication year. This strategy resulted in eleven sets of citations, one for each year.

### 3. Saving the search results

The sets of citations for each year were saved in the MEDLINE reprint format. This file format is commonly used for MEDLINE citations. Reprint format files consist entirely of plain ASCII text.

### 4. Importing citations into a reference management application

EndNote 3.0 was used for the reference management application. One EndNote library file was created for each of the eleven years 1988 to 1998. The citation files were then imported for each year into the appropriate EndNote library, until all of the citations were in EndNote.

EndNote was necessary as an intermediate step in the process of building the bibliographic database. EndNote libraries were designed to be used on one computer and could not be searched over the Internet. On the other hand, Microsoft Access, which *could* be connected to the Internet, was not able to read and

**Figure 1**  
OutcomeLit search page

The screenshot shows a web-based search interface. At the top, it says "OutcomeLit Search Form:". Below that, there is a date range selector: "From 19[88] to 1999." with a "Reset" button. Underneath is a search area with two text input fields. The first field contains the word "asthma" and the second field contains "emergency". Between these two fields are radio buttons for "AND" and "OR". At the bottom of the search area are two buttons: "Search now" and "Reset".

import MEDLINE citations. Therefore, EndNote was needed to (1) import the MEDLINE citations that were saved in the Reprint file format and (2) export the citations in a format recognizable to a database application like Microsoft Access. Using EndNote as an intermediate step also facilitated the process of updating the database (step 12, below).

### 5. Reformatting the citations

Prior to exporting the citations from EndNote, some reformatting was needed. First, EndNote's Change Text function was used to remove the carriage return characters between MeSH terms and replace them with semicolons. Second, periods and spaces that EndNote placed between authors' initials were removed by editing the Tab.ens export file format.

### 6. Exporting the citations as text files

EndNote has many different options for exporting citations. In this example, the libraries of citations were exported as tab-delimited text files. This file format was a standard one that used tab characters to mark where one field ended and the next field began. Later, when these text files were imported into the bibliographic database, the database program recognized those tabs as the demarcation characters between fields. Each of the eleven files of citations were exported in this manner.

### 7. Importing the citations into a database

Microsoft Access 97 was used as the database management application. A blank database was created and saved as the file "OutcomeLit.mdb." The 1988 citations that had been saved as a tab-delimited text file in step 6 were imported. This process was repeated for each of the ten other text files of citations, resulting in one Microsoft Access database, OutcomeLit.mdb, consisting of a single database table.

### 8. Installing the ODBC driver for the database

The goal was to make the OutcomeLit database searchable over the Web. This objective required installing the appropriate open database connectivity (ODBC) database driver. ODBC is a computer industry standard for client/server data access. It is a means of specifying data sources and the means by which their data can be accessed by other computers.

Because Microsoft Access 97 was used as the database application, the ODBC driver for Access 97 needed to be installed. Microsoft Windows NT Workstation comes with the ODBC driver for Access 97 already installed. Windows NT's ODBC Administrator program needs to be run, specifying that the ODBC driver for Access 97 should be connected to the OutcomeLit database. The ODBC Administrator program is located in the Control Panel of Windows NT.

Had the project needed to use an ODBC driver that was not already installed on the computer, two additional steps would have been necessary: (1) downloading the appropriate ODBC driver from the Internet and (2) installing the new ODBC driver onto the system. After performing these steps, the process would be as described above.

### 9. Creating a Web page front end

In order to make the bibliographic database accessible to clinicians, creation of a Web page front end with an easy-to-use interface was necessary. Two simple hypertext markup language (HTML) pages were created to provide this front end to the database. The first page was the OutcomeLit Home Page, which simply welcomed users to OutcomeLit and provided a link to the search page.

The search page (Figure 1) was modeled after the Internet Grateful Med search screen. There were two data entry areas and a button labeled Search Now. The user could type any string of text into the search form and then click the button to start the search. By de-

**Figure 2**  
HTML formatting

```
<html>
<head>
<title> Search Results </title>
<h2> OutcomeLit Search Results </h2> <hr> </head>
<body>
  <%begindetail%>
  <%author%>. <b><%title%></b>.
  <i><%journal%></i>. <%volume%>
  (<%number%>): <%pages%>, <%year%>. <br>
  <a href="./outcomeLit/abstract_page.idc?citation=<%citation_id%>">
  Abstract </a>
  <%enddetail%>
</body>
</html>
```

fault, two search terms were combined using a Boolean AND, although the user could click a button to combine terms with a Boolean OR instead. The search would retrieve all citations in the database that matched the search terms from 1988 forward. Users could easily change the starting year for the search.

## 10. Connecting the database to the Web front end

The HTML search forms described above simply sent the user's search terms to the Web server. At this point, a computer scripting language was needed. Two types of scripts needed to be written. The first script translated the search terms into a database query statement and was the reason the ODBC driver was installed for Access 97 (step 8, above). The ODBC driver received the query statement generated by the script and then searched the OutcomeLit.mdb database file.

After the database query was performed, the search results needed to be sent back to the user and displayed as a Web page. To do this procedure, the search results needed to be formatted in HTML. The second script performed this formatting (Figure 2). It received the results of the database query and embedded them in an HTML page. This HTML page was sent to the user, who then saw a list of citations (Figure 3).

Many different scripting languages are available for these tasks, such as Perl, JavaScript, and Microsoft's IDC. Microsoft IDC has been used because it is built into Peer Web Services, the Web server built in to Microsoft Windows NT Workstation. The scripts used for this project are available on the OutcomeLit Web site for anyone to copy or modify.

## 11. Publishing the bibliographic database

Finally, the bibliographic database needed to be published on the Web or on the institution's intranet. Peer Web Services was used for this application because it was built into Windows NT Workstation, although any Web server could be used. Publishing the database re-

**Figure 3**  
Query result page

**OutcomeLit Search Results**

---

Search terms:     **asthma AND emergency**

Bernard-Bonnin AC, Stachenko S, Bonin D, Charette C, Rousseau E. **Self-management teaching programs and morbidity of pediatric asthma: a meta-analysis**. *Journal of Allergy & Clinical Immunology*. 95(1 Pt 1): 34-41, 1995.  
Abstract

Black ER, Weiss KD, Erban S, Shulkin D. **Innovations in patient care: changing clinical practice and improving quality**. *Joint Commission Journal on Quality Improvement*. 21(8): 376-93, 1995.  
Abstract

Crain EF, Weiss KB, Fagan MJ. **Pediatric asthma care in US emergency departments. Current practice in the context of the National Institutes of Health guidelines**. *Archives of Pediatrics & Adolescent Medicine*. 149(8): 893-901, 1995.  
Abstract

quired three steps: (1) activating the Web server program, (2) placing the HTML files for the Web page front end in the appropriate directory on the server, and (3) setting the system security to permit remote user access to those Web pages.

## 12. Updating the database

The database consisted of citations from 1988 through 1998. The database was updated with new citations for 1999 using the following procedure. The MEDLINE search for outcome studies was rerun using the Ovid Windows client. The search results were then limited to English-language articles published in 1999. This resulted in 188 citations, which were saved in the MEDLINE Reprint file format. The resulting file, "updat99a.mdl," was then moved into a folder containing the EndNote libraries. EndNote was started and a new library, called "1999\_OutcomeLit.enl," was created. The updat99a.mdl citation file was imported into this new EndNote library. The citations were reformatted (step 5, above) and exported as a tab-delimited text file (step 6, above). Finally, Microsoft Access was started, and the text file was imported into the OutcomeLit.mdb database. Importing the file completed the update procedure.

Even after a calendar year is over, citations for articles published that year continue to appear in MEDLINE, necessitating an update of OutcomeLit for 1998. To do this updating, the procedure described above was used with one modification. Because 2,395 citations were already in the 1998 the EndNote library, prior to importing the new citations into the EndNote library for 1998, EndNote's import option was set to discard duplicate citations. This setting allowed the

**Table 1**  
Number of citations retrieved using various search terms

Search terms	Citations
Asthma AND emergency	113
Endarterectomy OR vascular surgery	64
Myocardial infarction AND beta	34
Disease management AND heart	7
Pancreatitis AND cost	3

importation of only the forty-five new citations for 1998. The duplicate citations were discarded.

These forty-five new citations were then copied and pasted into a temporary library. This temporary library contained only the forty-five new 1998 citations, while the updated 1998 library contained both the 2,395 old and the forty-five newly imported 1998 citations. The updated 1998 library was then closed. The forty-five new citations in the temporary EndNote library were reformatted and exported as a tab-delimited text file (step 5, above) and then imported into the OutcomeLit database. This process will need to be repeated periodically, until all of the citations for 1998 articles have been entered into MEDLINE.

### More detailed information

The information presented above should be sufficient for users familiar with the software applications involved. For those with less expertise, the OutcomeLit Web site,<sup>†</sup> maintained at the Donald A.B. Lindberg Information Center, contains highly detailed instructions, along with numerous screen shots of key steps in the process.

### IMPLEMENTATION

The OutcomeLit.mdb bibliographic database contained 13,803 citations of articles published between 1988 and 1999 and consisted of a single database table that occupied 33.6 MB of disk space. The computer system used was a 400-MHz Pentium II computer with 128 MB of RAM memory and the Windows NT 4.0 Workstation operating system.

Most searches of the entire database, using two search terms, took between eight and twelve seconds. Searches limited to one year took two to four seconds. Table 1 shows the number of citations retrieved from the entire database, using various search terms.

After documenting all of the steps to build the OutcomeLit database, it was deleted and the entire process initiated again and timed. Table 2 lists the steps of the process and the time required for each step. The update for 1999 citations took only twenty-

<sup>†</sup> The OutcomeLit site may be viewed at [www.hmi.missouri.edu/hmi/center/](http://www.hmi.missouri.edu/hmi/center/).

**Table 2**  
Time required for each step of the process

Task	Time
1. Searching MEDLINE with Ovid	
Running the search	6 minutes
Limiting search results by year	4 minutes
Saving the search results by year	35 minutes
2. Managing the citations with EndNote	
Importing citation files into EndNote	74 minutes
Formatting the citations	28 minutes
Exporting citation files from EndNote	12 minutes
3. Building the database	
Creating a new MS Access database	20 minutes
Importing citation files into Access	15 minutes
Total time to build the database	3 hours, 14 minutes

eight minutes. The update procedure took considerably less time than building the database because the number of citations was much smaller and because EndNote and Microsoft Access had already been configured.

The most time consuming part of the process was creating the hedge search strategy. Because writing the scripts to connect the database to the Web page search form could also be time consuming, scripts have been posted on the OutcomeLit Web site for others to copy or modify.

The total cost of the software used to build the OutcomeLit database was \$868.90 (Table 3). This total did not include the cost of the Ovid Windows client software. Some libraries have Microsoft Access because it is included in the professional version of Microsoft Office. Similarly, some libraries already have EndNote (or Reference Manager) or Windows NT Workstation. Because of these factors, the expense for software will vary, but it should not exceed \$868.90.

### DISCUSSION

There is nothing novel about using hedges to search MEDLINE, importing citations into EndNote, or making a database searchable over the Web. Nevertheless, linking these activities to create a Web-searchable bibliographic database is not obvious. As demonstrated in this study, the process can easily be done with common software tools and a few short computer scripts.

**Table 3**  
Cost

Software	Cost
Microsoft Windows NT Workstation 4.0	\$ 269.95
Microsoft Access 97	\$ 299.95
EndNote	\$ 299.00
Total	\$ 868.90

While Web-searchable databases have been described in the context of evidence-based health care, there are many other potential uses of this methodology, such as databases containing faculty bibliographies or information on grant-giving foundations and other funding sources.

The bibliographic database described lacks the sophisticated features that are available in tools like Ovid or PubMed and can never replace them. However, once a local bibliographic database has been built, enhancing its capabilities is relatively easy. Such enhancements fall into two categories: (1) multiple databases access and (2) improved search capabilities.

Once the basic system was up and running, a second database could be linked to the Web page front end. For example, due to a resurgence of tuberculosis, staff could create a TBLit database in the same way OutcomeLit was built. The process would be exactly the same, except that the scripts would send the structured query language (SQL) query statement to the TBLit database instead of the OutcomeLit database. The Web page front end would also have to be modified to ask users whether they wanted to search OutcomeLit or TBLit. In this way, many locally maintained bibliographic databases could be served through a single Web page interface.

The database search capabilities could also be improved once the basic system was operational. This improvement, too, would require new scripts and changes in the Web page search form. For example, to search for outcome studies concerning the treatment of asthma patients in the emergency department, the user could search on (1) "asthma AND emergency" and (2) "disease management OR pathway," and then combine those two searches using a Boolean AND. Many different enhancements of this sort could be added once the basic system has been built.

It should be noted that there are software applications that automate the process of writing scripts. For

example, Microsoft's Visual InterDev and Sybase's PowerDesigner use graphical tools (dialog boxes, drag-and-drop, etc.) to create the appropriate scripts. Using these tools, the user does not need to write scripts or even see them. For OutcomeLit, the scripts are short and easy to write and done without the use of scripting tools. Software applications like PowerDesigner or Visual Interdev would increase the expense and complexity of the methodology.

In describing the methodology for this study, specific software applications were used such as Ovid, EndNote, and Microsoft Access. These applications were chosen because they were inexpensive, easy to use, and widely available. However, the principles of this methodology were general, and the applications used could be replaced with other applications that provide similar functionality.

## CONCLUSION

With the growth of evidence-based health care, libraries are being called upon to provide up-to-date relevant literature on an increasing number of topics. The methodology described in this study allows librarians to compile the results of monthly SDI searches into a single Web-searchable database. Creating locally maintained bibliographic databases significantly extends the evidence-based health care services that libraries can offer to clinicians.

## REFERENCES

1. HARDIN LIBRARY FOR THE HEALTH SCIENCES. PubMed MEDLINE search. dermatology & skin diseases. [Web document]. Iowa City, IA: University of Iowa, 1999. [rev. 5 Apr 1999; cited 6 Dec 1999]. <<http://www.lib.uiowa.edu/hardin/pm/derm.html>>.
2. SIEVERT ME, BOYCE BR. Hedge trimming and the resurrection of the controlled vocabulary in online searching. *Online Rev* 1983 Dec;7(6):489-94.

*Received June 1999; accepted December 1999*