Steps in Planning and Implementing a Literature Search

Author
Barbara Folb, MM, MLS, MPH. Public Health Informationist / Librarian, University of Pittsburgh. folb@pitt.edu, 412-383-5165 or 412-648-1974.

Last Update
August 29, 2012

Table of Contents

Introduction.................................................................2
Step I: Define the Research Topic........................................4
Step II: Extract the Key Concepts from the Research Question.....5
Step III: Define Inclusion Criteria for Reviewing Articles..........5
Step IV: Do an Exploratory Search.....................................6
Step V: Search Improvement.............................................10
  Quick Search Approaches............................................19
Step VI: Save and Manage Searches and Citations..................23
Step VII: Get the Full Text Articles...................................25
Step VIII: Evaluate Articles............................................26
Step IX: Organize and Store Full Text Articles........................27
Glossary........................................................................29

Copyright Information:
This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.
INTRODUCTION

Definitions needed:

**Bibliographic Database** – A database describing journal articles, books, book chapters, dissertations, reports, and other written documents used as an aid to locating written materials on a topic. Traditionally a bibliographic database does not contain the full text of the documents, but in the online environment many bibliographic databases and collections of full text articles have been linked for easier movement from the citation to the actual document.

**Citation** - Data on a source of information that will allow identification of the original full text source. For example, for a journal article this would include author(s), article title, journal title, issue and volume; page numbers, year of publication, URL or doi number.

**PubMed** – PubMed is a bibliographic database containing citations to articles on all aspects of health, medicine, and biological sciences. It is produced by the National Library of Medicine at the National Institutes of Health, and is freely available on the Internet. It covers journal literature from to 1950s to the present. Links to many full text online journals are provided.

What this document covers

This document explains how to design, refine and manage the results of a scholarly literature search. The PubMed database is used, but the search steps can be applied to any database. Some discussion of syntax required by the PubMed database is included, but the focus is on aspects of searching that apply to any database.

Search Process Summary

The steps in the following list should be included in any search for an important project such as writing a term paper, journal article, thesis or dissertation, or research proposal. While you always start with the question, the remaining steps may flow in various orders and be repeated as needed until you have enough information to accomplish the writing task at hand.

- Define search question
- Extract the important concepts from the question
- Organize the important concepts into a search statement the database can interpret (i.e., use database syntax)
- Do the initial search
• Evaluate the citations retrieved for fit with search question
• Find and document alternative terms for the important concepts in the citations retrieved.
• Do more searching, testing alternative terms systematically
• Document the search process
• Save the good citations in a bibliographic management program such as EndNote or Mendeley.
• Retrieve the full text articles

Once you have the full text articles you are ready to move on from searching to evaluating and summarizing the information the articles present.

**What does a good search look like?**

There is no such thing as a perfect search. A good search is one that meets your information needs. For most people this means that the search:

• Retrieves more articles on your topic than ones that are off topic
• Retrieves a “reasonable” number of citations.

Searches are like laboratory tests, they have sensitivity and specificity. There will always be “false positives,” articles retrieved that are not on topic, and “false negatives,” good articles that are missing from the results. What is reasonable number of citations to review? That depends on how much time you have to evaluate the articles and what you will be doing with the articles found (10 page paper, dissertation, systematic review, etc).

**Real Searching can be Messy**

The search process is described in a very logical step-wise fashion here, but in real life searching is sometimes messy, especially when you are doing the initial exploratory searching. It is easy to get lost in the process by impulsively trying various strategies without a systematic plan. Be prepared to be creative and flexible in thinking about how concepts – abstract ideas – can be represented by words in many ways, but be organized in documenting what you try along the way. You will save time and energy if you document well, because you won’t keep trying the same things over and over again.

**What databases are important to Public Health Topics?**

A thorough search on a public health topic may require the use of databases from related fields such as sociology, anthropology, psychology, engineering, biology, and environmental sciences. For help in choosing databases based on topic, see the University Library System page “Databases by Subject” [http://www.library.pitt.edu/articles/dbsub.html](http://www.library.pitt.edu/articles/dbsub.html). There are also more
databases on health topics listed on the HSLS page “Databases A to Z http://www.hsls.pitt.edu/resources/databases”.

**STEP I: DEFINE THE RESEARCH TOPIC**

**Task:** State topic in one sentence

*A research topic that can be said with one sentence will usually contain enough concepts to define a good search.* If you have only a phrase you probably have not defined the topic well enough and will retrieve articles on divergent topics, unless that phrase is truly unique, and the topic very new. If it takes more than one sentence to define your topic you probably need to do several separate searches, or think about narrowing the scope of the question. A search using all the concepts from several sentences may retrieve nothing, or a small portion of the potential articles of interest.

The following list contains some of the most commonly used search parameters for biomedical and public health searches. Your search may not need all of them, but should include those that are critical to your topic.

- **Problem** (disease, risk factor, behavior to change, environmental condition, etc)
- **Population** (age, ethnicity, race, geography, etc)
- **Intervention** (education, treatment, environmental change, etc)
- **Setting** (school, home, community, clinic, country, rural or urban area, etc)
- **Service provider** (health educator, clinician, peers, etc)
- **Methods/ Theories of interest** Study design (ex: cohort study), theories (ex: Stages of change), methodologies (ex: qualitative or quantitative studies) of interest
- **Outcome(s) of interest** (reduced risk, acceptability of intervention, adoption of new behavior, decrease in incidence or prevalence, etc)

**Example of an initial search question:** *Does peer-to-peer education decrease risky sexual behavior in adolescents?*

Possible variations:

- focus on a more clearly defined behavior
include a narrower population defined by geography, ethnicity, sex, urban or rural setting, etc.

**STEP II: EXTRACT THE KEY CONCEPTS FROM THE RESEARCH QUESTION**

**Task:** Separate the search question into discrete concepts.

Most databases do not understand questions stated in sentence format. They expect to see separate concepts that are connected together with operators such as AND, OR and NOT. This allows them to process the terms in a logical manner. For more information on search operators in PubMed, see the *Introduction to Boolean Logic* in the PubMed online tutorial, [http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_350.html](http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_350.html).

The example question - *Does peer-to-peer education decrease risky sexual behavior in adolescents?* – can be separated out into the following concepts:

- Peer-to-peer education
- Risk reduction
- Sexual behavior
- Adolescents

Note that some topics can be combined or split. For example, “sexual risk taking” combines the 2nd and 3rd topics above. Whether it is best to separate topics into their smallest units, combine them into one concept, or use both approaches will vary from topic to topic.

**STEP III: DEFINE INCLUSION CRITERIA FOR REVIEWING ARTICLES**

**Task:** List the characteristics of articles that will help you build a logical argument in your literature review.

Before you search, it is helpful to decide what criteria will be used to decide which articles to keep for your literature review. Criteria beyond those defined in the search statement are usually needed. This many include information that is not in the PubMed record that can only be evaluated by reading the full text article. Commonly used criteria include:

- research design
- methodology and how well it is reported, including details such as
  - sample size
  - length of the study
  - theories used
• how closely the study population resembles the population you are interested in

Generally, the criteria are used in two ways. First they are used to decide which articles to get in full text and read. Second, they are used when reading the full text articles to evaluate their quality.

For large scale projects such as theses or systematic reviews it is helpful to set up a spreadsheet to organize and document the article review process. This will be discussed in more detail in Step VIII.

**STEP IV: DO AN EXPLORATORY SEARCH**

Task: Do a first search and harvest other possible search terms from it.

Definitions needed:

Concept – An abstract idea that can usually be represented more than one way with words and phrases. For example, the concept of “health promotion” could be referred to with the phrase “health campaign” or “wellness program” among others. Concepts can be expressed in the text of an article or abstract with words and phrases, or more formally defined with **controlled vocabulary** in a bibliographic database.

**Controlled Vocabulary** - A system of words and phrases used to standardized representation of concepts in a bibliographic database. Use of a controlled vocabulary allows the searcher to use one term or phrase and pull up articles on that topic that used other words to represent the concept. For example, in PubMed, an article that discusses Type II diabetes may use a number of phrases for the concept. The controlled vocabulary term “diabetes mellitus, type 2” will be assigned to all those articles.

**MeSH** - Acronym for Medical Subject Headings. The controlled vocabulary used to index citations in PubMed.

Creating the Initial Search, Starting the Documentation

One important goal for the first search is to discover what words and phrases are used to represent concepts of interest in the title, abstract, or indexing of the articles that most closely match your topic. For example, the concept “risk” could be represented by “risk taking,” or “risky behavior,” or by specific types of behavior such as unprotected sex or its opposite, condom use. Finding and testing these words and phrases is an extremely important step in creating and refining a literature search that meets the information need at hand.
Take the main concepts extracted from the research question in Step III and enter them into a table that you can use to track your search progress. Here is an example of a search tracking table based on the question from step III. This is what it looks like before doing the first search:

<table>
<thead>
<tr>
<th>Peer-to-peer education</th>
<th>Risk reduction</th>
<th>Sexual behavior</th>
<th>Adolescents</th>
</tr>
</thead>
</table>

Each column contains one concept. The AND operator will be used to connect the concepts in the PubMed search. You can either enter them all in one string in the search box, like this:

*Peer to peer education AND risk reduction AND sexual behavior AND adolescents*

Or, you can enter them one at a time in PubMed, do the search, and combine them on PubMed’s Advanced page using the Search History. The advantage to the second method, which is initially more work, is you can reuse and recombine your concepts without having to reenter them all. For more information on using the PubMed Search History see [http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_400.html](http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_400.html).

Here is a screen shot of a PubMed search history, showing both methods of the initial search:

<table>
<thead>
<tr>
<th>Search</th>
<th>Add to builder</th>
<th>Query</th>
<th>Items found</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>#6</td>
<td>Add</td>
<td>Search #4 AND #5 AND #6 AND #7</td>
<td>90</td>
<td>14:21:12</td>
</tr>
<tr>
<td>#7</td>
<td>Add</td>
<td>Search adolescents</td>
<td>1455022</td>
<td>14:36:24</td>
</tr>
<tr>
<td>#5</td>
<td>Add</td>
<td>Search sexual behavior</td>
<td>26736</td>
<td>14:36:18</td>
</tr>
<tr>
<td>#5</td>
<td>Add</td>
<td>Search risk reduction</td>
<td>166510</td>
<td>14:36:18</td>
</tr>
<tr>
<td>#4</td>
<td>Add</td>
<td>Search Peer to peer education</td>
<td>12211</td>
<td>14:29:57</td>
</tr>
<tr>
<td>#3</td>
<td>Add</td>
<td>Search Peer to peer education AND risk reduction AND sexual behavior AND adolescents</td>
<td>90</td>
<td>14:29:41</td>
</tr>
</tbody>
</table>

When run on 1/3/2012, both versions of the search retrieved 90 citations, because they are exactly the same, just entered differently.

- The line labeled #3 shows the search entered all at one time.
- The lines labeled #4- #7 show the concepts entered one at a time.
- The line labeled #8 shows the combined statement including lines #4-#7.

Looking at the results, it becomes immediately apparent that studies from all over the world (Tanzania, United States, Uganda, etc) and from various demographic groups (Mexican Americans, US Pacific Islanders, youth in foster care, etc) have been retrieved, and that application of demographic limits may
be needed at a later point in the search. Only a few citations are on the correct topic, but almost every citation contained useful terms to consider for further searching, so it is still useful to look through them at this point.

In addition to harvesting the terms for future use, save any articles you of interest to the PubMed Clipboard using the “Send to” drop down menu. When you are done with the search session you will want to save the Clipboard contents to a MyNCBI collection for future use. For more on using the PubMed Clipboard see http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/030_220.html. For more on using MyNCBI collections see http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/030_250.html

Examining Results, Harvesting New Search Terms

You will fill in the search tracking table as you evaluate the search. In the cell beneath each concept enter synonyms and closely related terms you discover and decide to include. Synonyms can be harvested from the MeSH (Medical Subject Headings), the article title, and the abstract.

To display the entire record for each citation in PubMed, use the Display Settings drop down box and under Format choose Abstract:

Here is an excerpt from an abstract retrieved in the search. A synonym for peer-to-peer is highlighted:

ethnicity. The results of the economic evaluation were considered to be illustrative, mainly due to the uncertainty of the effect of intervention on behavioural outcomes. The results were most sensitive to changes in parameter values for the intervention effect, the transmission probability of STIs and the number of sexual partners. The costs of teacher-led and peer-led interventions, based on the resources estimated from the relevant randomised controlled trials in our systematic review, were 4.30 pounds and 15 pounds per pupil, respectively. Teacher-led interventions were more cost-effective than peer-led interventions due to the less frequent need for training. The incremental cost-effectiveness of the teacher-led and peer-led interventions was 20,223 pounds and 80,782 pounds per quality-adjusted life-year gained, respectively. An analysis of individual parameters revealed that future research funding should focus on assessing the intervention effect for condom use from a school-based intervention.
That term is entered into the search table for later testing:

<table>
<thead>
<tr>
<th>Peer-to-peer education</th>
<th>Risk reduction</th>
<th>Sexual behavior</th>
<th>Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Peer-led”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Looking at the MeSH headings, no MeSH heading for peer-to-peer education exists. However, there are MeSH terms for some of the other concepts, as illustrated by the list of MeSH headings attached to one of the citations:

**MeSH Terms:**
- Adolescent
- Adolescent Behavior*
- Cost-Benefit Analysis
- Female
- Great Britain
- Humans
- Male
- Risk Reduction Behavior*
- Safe Sex
- Schools
- Sex Education
- Sexually Transmitted Diseases/prevention & control*
- Young Adult

Here is the concept table, after adding the appropriate MeSH headings from the list above. Words and phrases followed by a / are MeSH terms. Words in “ “ appeared in titles or abstracts. This helps to clearly show which are text words and which are MeSH headings:

<table>
<thead>
<tr>
<th>Peer-to-peer education</th>
<th>Risk reduction</th>
<th>Sexual behavior</th>
<th>Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Peer-led”</td>
<td>Risk reduction behavior/</td>
<td>Safe sex/</td>
<td>Adolescent/</td>
</tr>
<tr>
<td></td>
<td>Sexually transmitted diseases/prevention &amp; control</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use of synonyms is especially important when looking for concepts not included in the MeSH controlled vocabulary. As noted above, there is no
MeSH heading for peer to peer education. In the following table, under peer to peer education, more phrases that were used by authors to represent the concept are listed. Words and phrases followed by a / are MeSH terms. Words in “ “ appeared in titles or abstracts.

<table>
<thead>
<tr>
<th>Peer-to-peer education</th>
<th>Risk reduction</th>
<th>Sexual behavior</th>
<th>Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Peer-led”</td>
<td>Risk reduction</td>
<td>Safe sex/</td>
<td>Adolescent/</td>
</tr>
<tr>
<td>“peer education”</td>
<td>behavior/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“peer-based”</td>
<td>Sexually</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“peer educator”</td>
<td>transmitted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“peer delivered”</td>
<td>diseases/prevention &amp; control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“peer informants”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above could be expanded, adding more terms to columns 2-4, but there is enough information to evaluate how useful the peer-to-peer synonyms as a demonstration of the process. How far you go in adding and refining as search is entirely decided by you. When you have enough articles on your topic, you are done, no matter what the table looks like.

**STEP V: SEARCH IMPROVEMENT**

**Tasks:**
- Test the terms harvested from the initial search one by one to see if adding them to the search improves the results.
- Learn techniques for addressing common search problems such as retrieving too many or too few citations, or too many off topic citations.

**Testing of Harvested Terms**

**Critical point:** terms must be tested one at a time in order to determine their effect on citation retrieval.

To test the peer-to-peer synonyms, first create a search that includes all the other pieces of the initial search. In the following screen shot, this has been done reusing search statements from the History (#9):
This is the equivalent of typing in the search

- risk reduction AND sexual behavior AND adolescents.

Before entering the new synonyms to test, test a variation on peer to peer education by putting it in quotation marks, indicating that you want it to be a phrase. Note that in the following screen shot, it gets a much smaller number of citations with the quotation marks (#10) than it does without the quotation marks (#4). In fact, when combined with the rest of the search (#11) “peer to peer education” retrieves no citations:

Now test the term “peer led”:

Note that “peer led” retrieves many more citations on its own (#12) than “peer to peer education” (#10). When combined with the rest of the search, “peer led” retrieves 8 citations (#13).
All 8 citations retrieved look potentially useful by their titles (box below), reading the abstracts will help decide which to keep. At this point, any that look useful should be saved temporarily to the PubMed clipboard, and later to the PubMed Collections for future use.


Based on this search, it is reasonable to keep “peer led” as part of the search. It is also reasonable to test “peer to peer” without the word education (#14). When combined with the rest of the search, it does retrieve one citation (#15), but it is not on topic. However, since it doesn’t retrieve a lot of off topic citations that would muddy up the search, it can either be kept or discarded without much effect.

The next step is to continue testing the text word synonyms for peer to peer. **One important step is added, examining the citations that are retrieved by the new term that were not retrieved by the old term.** Removing the duplicate citations both cuts down on the time needed to evaluate the search terms and allows you to see exactly what the new term adds to the retrieval. This is critical to deciding whether the new term is adding relevant citations to the retrieval. It should be done for every new term after it is added.

To isolate the new citations retrieved, use the NOT operator. The general form looks like this:

- new search NOT old search

Here is an example spelled out with all the terms:

**NEW SEARCH:** “peer education” AND risk reduction AND sexual behavior AND adolescents

**OLD SEARCH:** “peer led” AND risk reduction AND sexual behavior AND adolescents

**SEARCHES COMBINED WITH NOT:** (“peer education” AND risk reduction AND sexual behavior AND adolescents) NOT (“peer led” AND risk reduction AND sexual behavior AND adolescents)

**Note the parentheses** around each search section in the combined search. These tell the search engine to run what is inside each set of parentheses and then combine them. **Without the parentheses, PubMed runs the search left to right, and you may get different results, depending on the order of the operators.**
The combined search above does not have to be re-entered as it is above. Instead, existing search lines in the Search History can be combined. To do this, on the Advanced page:

- Left click on the #number to the left of the new search, then click on AND in the drop down box. The search is added to the Search Box.
- Left click on the #number in front of the old search, then click on NOT. The search is added to the Search Box, with the ( ) needed.
- Click Preview to see how many are retrieved.
- Click on the number of citations retrieved to see the unique citations from the new search not retrieved by the old search.

Here is screen shot showing the results. #17 is the new search, which gets 14 citations. Thirteen of those are not in the old search, as shown in #18:

<table>
<thead>
<tr>
<th>Search</th>
<th>Add to builder</th>
<th>Query</th>
<th>Items found</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>#18</td>
<td>Add</td>
<td>Search #17 NOT #13</td>
<td>13</td>
<td>15:52:46</td>
</tr>
<tr>
<td>#17</td>
<td>Add</td>
<td>Search #5 AND #16</td>
<td>14</td>
<td>15:51:29</td>
</tr>
<tr>
<td>#16</td>
<td>Add</td>
<td>Search &quot;peer education&quot;</td>
<td>456</td>
<td>15:51:16</td>
</tr>
<tr>
<td>#15</td>
<td>Add</td>
<td>Search #5 AND #14</td>
<td>1</td>
<td>14:52:41</td>
</tr>
<tr>
<td>#14</td>
<td>Add</td>
<td>Search &quot;peer to peer&quot;</td>
<td>235</td>
<td>14:52:25</td>
</tr>
<tr>
<td>#13</td>
<td>Add</td>
<td>Search #9 AND #12</td>
<td>2</td>
<td>14:51:49</td>
</tr>
<tr>
<td>#12</td>
<td>Add</td>
<td>Search (&quot;peer led&quot;)</td>
<td>316</td>
<td>14:49:51</td>
</tr>
<tr>
<td>#11</td>
<td>Add</td>
<td>Search #9 AND #10</td>
<td>0</td>
<td>14:47:32</td>
</tr>
<tr>
<td>#10</td>
<td>Add</td>
<td>Search &quot;peer to peer education&quot;</td>
<td>10</td>
<td>14:47:16</td>
</tr>
<tr>
<td>#9</td>
<td>Add</td>
<td>Search #5 AND #6 AND #7</td>
<td>174</td>
<td>14:44:42</td>
</tr>
<tr>
<td>#8</td>
<td>Add</td>
<td>Search #4 AND #5 AND #6 AND #7</td>
<td>90</td>
<td>14:41:36</td>
</tr>
<tr>
<td>#7</td>
<td>Add</td>
<td>Search adolescents</td>
<td>145823</td>
<td>14:36:24</td>
</tr>
<tr>
<td>#6</td>
<td>Add</td>
<td>Search sexual behavior</td>
<td>147689</td>
<td>14:36:16</td>
</tr>
<tr>
<td>#5</td>
<td>Add</td>
<td>Search risk reduction</td>
<td>165510</td>
<td>14:36:19</td>
</tr>
<tr>
<td>#4</td>
<td>Add</td>
<td>Search Peer to peer education</td>
<td>112911</td>
<td>14:26:57</td>
</tr>
<tr>
<td>#3</td>
<td>Add</td>
<td>Search Peer to peer education AND risk reduction AND sexual behavior AND adolescents</td>
<td>90</td>
<td>14:24:51</td>
</tr>
</tbody>
</table>

The next step is to look at the unique citations from the new search, and decide whether they are mostly on or off subject, and whether to keep them. Below are the 13 unique citations. Most are related to the topic, and the search term tested should be kept.


Each time a new term is evaluated, and you decide to keep it in the search, a new search statement should be saved that has all the terms that will be kept. Here is the search statement so far:

- (“peer led” OR “peer education”) AND risk reduction AND sexual behavior AND adolescents
This is what it looks like in the search history (#20):

<table>
<thead>
<tr>
<th>Search</th>
<th>Add to builder</th>
<th>Query</th>
<th>Items found</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>#20</td>
<td>Add</td>
<td>Search #9 AND #19</td>
<td>21</td>
<td>15:46:09</td>
</tr>
<tr>
<td>#12</td>
<td>Add</td>
<td>Search #12 OR #16</td>
<td>762</td>
<td>15:46:46</td>
</tr>
<tr>
<td>#13</td>
<td>Add</td>
<td>Search #17 NOT #13</td>
<td>13</td>
<td>15:44:34</td>
</tr>
<tr>
<td>#17</td>
<td>Add</td>
<td>Search #9 AND #16</td>
<td>14</td>
<td>15:42:14</td>
</tr>
<tr>
<td>#16</td>
<td>Add</td>
<td>Search &quot;peer education&quot;</td>
<td>466</td>
<td>15:42:09</td>
</tr>
<tr>
<td>#15</td>
<td>Add</td>
<td>Search #9 AND #14</td>
<td>1</td>
<td>14:52:41</td>
</tr>
<tr>
<td>#11</td>
<td>Add</td>
<td>Search &quot;peer to peer&quot;</td>
<td>235</td>
<td>14:52:25</td>
</tr>
<tr>
<td>#13</td>
<td>Add</td>
<td>Search #9 AND #12</td>
<td>0</td>
<td>14:51:49</td>
</tr>
<tr>
<td>#12</td>
<td>Add</td>
<td>Search &quot;peer led&quot;</td>
<td>315</td>
<td>14:49:51</td>
</tr>
<tr>
<td>#11</td>
<td>Add</td>
<td>Search #9 AND #10</td>
<td>0</td>
<td>14:47:32</td>
</tr>
<tr>
<td>#10</td>
<td>Add</td>
<td>Search &quot;peer to peer education&quot;</td>
<td>10</td>
<td>14:47:16</td>
</tr>
<tr>
<td>#9</td>
<td>Add</td>
<td>Search #5 AND #6 AND #7</td>
<td>1210</td>
<td>14:44:42</td>
</tr>
<tr>
<td>#8</td>
<td>Add</td>
<td>Search #4 AND #6 AND #7</td>
<td>90</td>
<td>14:41:36</td>
</tr>
<tr>
<td>#7</td>
<td>Add</td>
<td>Search adolescents</td>
<td>1458623</td>
<td>14:39:24</td>
</tr>
<tr>
<td>#5</td>
<td>Add</td>
<td>Search sexual behavior</td>
<td>96786</td>
<td>14:30:16</td>
</tr>
<tr>
<td>#5</td>
<td>Add</td>
<td>Search risk reduction</td>
<td>185510</td>
<td>14:30:19</td>
</tr>
<tr>
<td>#4</td>
<td>Add</td>
<td>Search Peer to peer education</td>
<td>11211</td>
<td>14:29:57</td>
</tr>
<tr>
<td>#2</td>
<td>Add</td>
<td>Search Peer to peer education AND risk reduction AND sexual behavior AND adolescents</td>
<td>50</td>
<td>14:29:41</td>
</tr>
</tbody>
</table>

Now let’s skip ahead, and show the search history after all the peer to peer synonyms have been tested. All of the yellow highlighted terms are included in the peer to peer concept of the search. When combined with #9, the as yet untested section of the search, there are 29 citations.:
The next step is to complete the harvesting of the synonyms for the other terms and test those the same way the peer to peer terms were tested.

Don’t forget to continue to save any articles you of interest to the PubMed Clipboard using the “Send to” drop down menu.
### The Search Table after all MeSH Terms are Harvested

The table now has all the MeSH terms of interest from the 29 citations in the final set (#40) above. This could be expanded by also including text words in the title and abstract, but only if needed to find more articles.

<table>
<thead>
<tr>
<th>Peer-to-peer education</th>
<th>Risk reduction</th>
<th>Sexual behavior</th>
<th>Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Peer-led”</td>
<td>Risk reduction behavior/</td>
<td>Safe sex/</td>
<td>Adolescent/</td>
</tr>
<tr>
<td>“peer education”</td>
<td>Sexually transmitted diseases/prevention &amp; control</td>
<td>Condoms, Female/utilization</td>
<td></td>
</tr>
<tr>
<td>“peer-based”</td>
<td>HIV Infections/prevention &amp; control</td>
<td>Unsafe Sex/</td>
<td></td>
</tr>
<tr>
<td>“peer educator”</td>
<td>Risk-Taking/</td>
<td>Sexual Behavior/</td>
<td></td>
</tr>
<tr>
<td>“peer delivered”</td>
<td>Health Knowledge, Attitudes, Practice/</td>
<td>Sexual Partners/</td>
<td></td>
</tr>
<tr>
<td>“peer informants”</td>
<td>Health Behavior/</td>
<td>Condoms/</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contraception/</td>
<td></td>
</tr>
</tbody>
</table>

To test the MeSH terms, you need to search them with the field tag [MeSH]. Otherwise, the search looks for the words anywhere in the citation. You can use the Search Builder on the Advanced page to search the MeSH, which will ensure that the syntax is correct, as shown in the next screenshot:
Clicking on the term you want, then on the Search button will launch the search.

**If you are building a complex search, you should save each concept search in MyNCBI before moving on to testing the next concept.** That way, if you exceed the PubMed limit of 100 search statements your work will not disappear from the Search History. **See Step VI for more information on saving searches.** After saving the peer to peer search, you would test the risk reduction terms using the same process that you used for the peer to peer terms.

Once you complete testing of all the concepts, you can consider whether or not the topic would benefit from further limitation, possibly by geography, race, gender, or setting of the intervention.

**Quick Search Approaches**

For many class writing assignments, the meticulous approach shown above may be too much work for the relative importance of the project. When something quicker is needed, consider the following approaches:

**Related Citations Searching**

The PubMed Related Citations search uses an article you select as a model, and looks for the other articles in PubMed that most closely resemble your article. It will return around 100 citations, relevancy ranked, meaning the ones that most closely match your citation will appear first.

The first few Related Citations appear automatically at the top right of the search results page. Click on the “see all” link to view the rest of them.

**Example Article:**


**Related citations retrieved by PubMed based on Mahat article (1st 10 shown):**


**Google Scholar**


Google Scholar defaults to a relevancy ranked set of search results which can easily be reviewed, but the opportunities for improving a search sequentially are not provided. Try using your best terms for each concept in Google Scholar which doesn’t allow for ORing a lot of synonyms together.

### Addressing Common Search Problems

**I didn’t find anything on my topic.**

- **Look closely at your search syntax.** Putting in the wrong operator, not using parentheses around OR’d synonyms, or failing to capitalize an operator in a database that requires it can really change your results. Also double check your spelling. Some databases such as PubMed try to suggest alternatives to misspelled words, others just retrieve citations that also misspell the words.
Consider using different words to represent your concepts. There may be different terms that is used more frequently, or an alternative spelling to consider.

Remove the least important search concept. If you have defined the topic very narrowly, it may not retrieve anything.

Try a different database. Perhaps you are looking in the wrong place. For example, a question related to a business aspect of health care will be more fully covered in a business database such as Ebsco Business Source Premier than it will in MEDLINE.

When was most of the research done on this topic? It is possible that your topic is either very new, or very old. If it is very new, consider looking into databases such as NIH's RePORTER database - http://projectreporter.nih.gov/reporter.cfm of funded research projects, or the Health Services Research Projects in Progress database - http://wwwcf.nlm.nih.gov/hsr_project/home_proj.cfm - or conference proceedings. There you can identify research that may not be published yet but is in the pipeline. If your topic includes a lot of research done pre 1966, it may not be well represented in the online databases. Use older print indexes, or retrospective databases such as JSTOR or the older Index Medicus records that have been added to MEDLINE. Keep in mind that different terminology may have been used when the research was first being done.

I found a few good articles, but not very many.

Build on those few good articles. Extract other possible search terms from the title, abstract, and indexing, and or them in to your search. Try using the PubMed Related Citations feature to find other articles that closely resemble the ones you have.

See if anyone has cited those few good articles. Databases such as Web of Science and Scopus, available from the HSLS Databases A to Z page, http://www.hsls.pitt.edu/resources/databases allow you to see who has cited an article. Of course this works best if the article you have is a few years old.

Try Google Scholar to see if its algorithm brings anything new to the surface. Google Scholar sorts its search based on a relevancy ranking algorithm. It may bring relevant articles to the top of the list that you can then go look for in PubMed and see if they provide clues on how to get more articles of interest out of the PubMed search. Google Scholar is a great tool, but should be supplemented by searches in other literature
I found too many articles on my topic.

**Limit by time**, such as last 10 years. This is a quick and easy way to cut down on the amount, but it is not usually the best choice as you may miss relevant older articles. For topics with a definitive start date, (ex: SARS) the literature search itself will naturally create that cutoff date if you have defined the topic clearly.

**Add more search criteria.** This makes sense if you can more specifically define your search. Adding one or two well-chosen concepts can make a significant dent in the amount of citations.

**Search for terms in specific fields.** For example, if your concept appears in the title of the article, or in a Major MeSH heading assigned to the article, it is likely to be more on topic.

**Skim the first 50-100 articles for ideas on how to improve the search.** If a few really good articles are there, examine their indexing and vocabulary used in the title and abstract for ideas on how to represent you topic.

I want reports of randomized controlled trials, but mostly I got case reports and case control studies.

**Use a search filter.** A variety of search filters have been created that are designed to increase the number of citations retrieved that utilize a specific research design, such as randomized controlled trials. Some of them have been validated. Filters vary in their quality. For a good overview of what a filter is, and links to some existing filters see [http://www.bestbets.org/links/search-strategies.php](http://www.bestbets.org/links/search-strategies.php). The article by Zhang et al “Optimizing search strategies to identify randomized controlled trials in MEDLINE” - [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1488863](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1488863) – gives information on the testing of changes to a widely used filter for retrieval of RCTs.

**RCTs on your topic may not exist.** If using a filter doesn’t retrieve RCTs, consider moving to other types of studies appropriate for your topic. See the Centre for Evidence Based Medicine (CEBM) levels of evidence information at [http://www.cebm.net/index.aspx?o=1025](http://www.cebm.net/index.aspx?o=1025).
**STEP VI: SAVE AND MANAGE SEARCHES AND CITATIONS**

**Task:** Save and organize your search strategies and citations in electronic and/or print form systematically.

**Saving Search Strategies**

Keeping track of your search activity can help you avoid wasting time on searches you have already tried. It also allows you to document your searches for others. If you save your searches in your MyNCBI account and/or copy them to a spreadsheet, you can easily track what you have already tried. MyNCBI also allows you to rerun the searches on demand or to set up periodic emails with new citations retrieved by your saved search.

Saving the strategy is very important if you are working on a long term project such as a dissertation or systematic review. Between the finalization of your search strategy and the analysis the citations and writing of the review time may elapse in which other appropriate articles could be published. You will need to rerun the search and update the list of citations.

Published systematic reviews of the literature require reporting of the search strategy as part of the methodology. The strategy used, the databases searched, the date(s) the searches were run, and the year span of literature searched are among the information that may be requested. If you are required to document your search for a project, you can distill a complex search down to one statement if needed. For good examples of search documentation, see articles published in the Cochrane Database of Systematic Reviews, which HSLS subscribes to.

It is easy to copy and paste a search history from PubMed or other databases into a spreadsheet such as Excel or a word processing document. The spreadsheet may not automatically recognize the column and row structure of the history, and copy it correctly. If it doesn’t, in Excel, try using Paste Special > Unicode. You can create a different worksheet for each database searched, and collect them all on one spreadsheet file.

**Saving Citations**

Three options for saving citations are discussed here, MyNCBI Collections, EndNote, and Mendeley.

**My NCBI Collections**

PubMed allows you to save sets of citations in MyNCBI in the Collections section. This is a good place to accumulate articles while you are in the midst of the search process. MyNCBI Collections are permanent. They will not
disappear unless you delete them. This is different than the PubMed Clipboard where you can accumulate citations during a search session and access them on your computer for up to 8 hours. Then they are deleted. Once you feel you have the citations you need, it is a good idea to also save those citations in EndNote or RefWorks which both make it easier to create bibliographies and references in papers.

**EndNote**

EndNote allows you to create databases to store and annotate your citations. The software can be used in conjunction with Microsoft Word to create and format a paper’s bibliography. EndNote is available on all public computers in the Falk Library and the University computing labs around campus. It can be purchased from Software Licensing Services (105 Bellefield Hall) by University of Pittsburgh students at a much reduced price.


When you are ready to move the citations from PubMed to EndNote, go to your MyNCBI account, and open the Collections tab. Click on the name of the collection, and it will display all the citations. Next, you must use the PubMed “Send to” menu to save the citations. Choose the “Citation Manager” option under Choose Destination. Then follow the browser file save prompts. For further instructions on creating the file see [http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/030_150.html](http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/030_150.html).

After you have saved the file, follow EndNote’s instructions on importing a file to add the citations to the EndNote database.

You can choose to create a new EndNote database for each project or one big database that holds all your collected citations. When you decide which articles to include in your review, it may be useful to copy those articles to a new EndNote database, or to add the information to the existing EndNote record so that you can easily identify that set of articles.

**Mendeley**

Mendeley is a bibliographic management program with two components, a desktop software program that can be downloaded to every computer that you work with, and a cloud computing version that can sync with your desktop databases and transfer files between the Internet and your various computers.
Mendeley is available at no cost. The University of Pittsburgh is in the process (as of 8/29/2012) of rolling out a subscription to Mendeley that will give University faculty staff and students more storage space in Mendeley.

Mendeley differs from EndNote in the way it imports article records into the database. Mendeley is set up assuming that users want to grab individual citations or pdfs of papers on the web, and use those to build the database. It does not allow for the easy importing of more than one citation from a database such as PubMed at this time, and so cannot be recommended for extensive projects.

Like EndNote, Mendeley allows users to insert citations into a paper written in Word on either a Mac or Windows computer. It assists with formatting of citations with in the text and in the reference list for the paper.

**STEP VII: GET THE FULL TEXT ARTICLES**

**Task:** Determine if the articles are available full text online, in print, or through HSLS Document Delivery.

There are a few general principles to keep in mind when moving from the citation to the full text of an article. These are very important, and may save you a lot of time.

1. No bibliographic database, including PubMed, links to all the full text journals that HSLS licenses for your use.

2. PittCat, [http://pittcat.hsls.pitt.edu](http://pittcat.hsls.pitt.edu), the online catalog, is the most complete source of information on which journals we have, whether in print or online. Always consult it if the links to an article are missing or broken.

3. Most electronic journals are available from more than one vendor. If HSLS doesn’t have it, the University Library System (ULS) may have purchased it. It is common to find journals available at Pitt from 2-6 providers with varying years of coverage. PittCat will show you this information.

4. For journals that we do not own, the HSLS Document Delivery Department can supply you with a copy at a subsidized rate. See [http://www.hsls.pitt.edu/services/documentdelivery](http://www.hsls.pitt.edu/services/documentdelivery) for details.

**Linking to Fulltext Articles from PubMed**

- View the Abstract display of the article you want
• Look for one or more links:
  o Links@Pitt-UPMC button will take you to some full-text provided through an HSLS subscription. NOTE: this button is not 100% accurate. If it doesn't provide a viable option, try the publisher link button in PubMed; or go to the journal through PittCat; or the use the journal location tools on the HSLS website at http://www.hsls.pitt.edu/resources/journals/.
  o A free full-text link will take you to the full-text provided free by the journal publisher.
  o A publisher full-text link will take you to the full-text if HSLS has a subscription to the journal. If HSLS does not have a subscription, you will be told that the article cannot be viewed without a subscription or payment.
• VERY IMPORTANT: If there are no full-text link options try PittCat anyway just to be sure.

Using PittCat to Link to Fulltext Online Journals
• In the Search Term(s): box enter the exact title, or at least the first few words, of the journal title
• In the Search Type: box select Journal Title Begins with
• Click the Search button
• Results of the PittCat search will tell you if the journal is available at any Pitt library in electronic or print format.
  o Electronic journals are noted with the words [Electronic Resource] after the title. Click on the title to see which years and volumes are available, and to link directly to the online journal.
• If the journal is not available at any Pitt library, the PittCat screen will read PittCat found no records that matched your search

STEP VIII: EVALUATE ARTICLES

Task: Using the criteria determined in Step II, evaluate full text articles for inclusion or exclusion from your review. Create a system for recording notes and decisions on each article.

One way to organize this process is to use a spreadsheet to enter the information you extract from each article. Set up a column for each of the review criteria. Record one article per row. Below is an example of part of such a spreadsheet. Other columns with further criteria would be in the actual spreadsheet. There would be a column noting whether the article would be included in the review.
### Steps in Planning and Implementing a Literature Search

**Lit Review, Peer to Peer Education Search**

<table>
<thead>
<tr>
<th>Article</th>
<th>Population</th>
<th>Setting</th>
<th>Intervention</th>
<th>Sample Size</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Health Organ Tech Rep Ser. 2006;938:103-50;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Step IX: Organize and Store Full Text Articles

**Concepts:**

- **Copyright** - the legal right of a copyright holder of a work such as a book, journal, or other material covered by copyright to control copying, distribution, and performance of the copyrighted material.

- **Fair Use** – Exceptions made to copyright law to allow for education and critical use of copyrighted materials. See [http://www.library.pitt.edu/guides/copyright/](http://www.library.pitt.edu/guides/copyright/) for more information.

**Task:** Set up a filing system for electronic and/or print article copies that works for you, and complies with copyright law.

If you save electronic copies, use a consistent naming style. I use `YearJournalAbbreviationAuthor` resulting in names such as `2006JKnowMngBhardwaj.pdf`. This allows for sorting the file folder by date and then by journal, and gives a clue as to which article is in the file. Record the location and name of the file in your bibliographic database so you can quickly retrieve the article.

If you save print copies, be consistent in how you file them, for example, by author last name then by year.

Copyright law governs how you can use someone else’s intellectual property. Under fair use in an educational setting you may have a copy of the article for
personal use without paying copyright fees. You may not keep multiple copies, or give out copies to others. Copyright law applies to all the journals that HSLS buys in print format.

Copyright law is complex. The University of Pittsburgh provides copyright guidelines that you may view at http://www.library.pitt.edu/guides/copyright/.

In the case of electronic full text journals, the licensing agreement with the provider overrides copyright law. In most cases it allows for similar use by students as copyright does. For example, you may not redistribute the article to others outside of the University, and you may not post an electronic copy in a publicly accessible place such as a website, even though that is easy to do. For any questions regarding allowable use of HSLS licensed journal content, please contact HSLS Technical Services at 412-648-9606.
GLOSSARY

**Bibliographic Database**
A database describing journal articles, books, book chapters, dissertations, reports, and other written documents used as an aid to locating written materials on a topic. Traditionally a bibliographic database does not contain the full text of the documents, but in the online environment many bibliographic databases and collections of full text articles have been linked for easier movement from the citation to the actual document.

**Citation**
Data on a source of information that will allow identification of the original full text source. For example, for a journal article this would include author(s), article title, journal title, issue and volume; page numbers, year of publication, URL or doi number.

**Concept**
An abstract idea that can usually be represented more than one way with words and phrases. For example, the concept of “health promotion” could be referred to with “health campaign” or “wellness program” and so forth. Concepts can be expressed with **words, terms, text words, subject headings, controlled vocabulary**, and so on.

**Controlled Vocabulary**
A system of words and phrases used to standardized representation of concepts in a bibliographic database. Use of a controlled vocabulary allows the searcher to use one term or phrase and pull up articles on that topic that used other words to represent the concept.

**Copyright** - the legal right of a copyright holder of a work such as a book, journal, or other material covered by copyright to control copying, distribution, and performance of the copyrighted material.

**Fair Use** – Exceptions made to copyright law to allow for education and critical use of copyrighted materials. See [http://www.library.pitt.edu/guides/copyright/](http://www.library.pitt.edu/guides/copyright/) for more information.

**Health Sciences Library System (HSLS)**
HSLS includes the main medical library, Falk Library, located in Scaife Hall. This is the main libraries used by public health students at the University of Pittsburgh.
**MeSH**
Acronym for *Medical Subject Headings*. The controlled vocabulary used to index citations in PubMed.

**PubMed**
PubMed is a bibliographic database containing citations to articles on all aspects of health, medicine, and biological sciences. It is produced by the National Library of Medicine at the National Institutes of Health, and is freely available on the Internet. It covers journal literature from to 1900s to the present. Links to many full text online journals are provided.

**Systematic Review**
A key tool to evidence based practice in many fields, a systematic review of the literature on a topic strives to retrieve and evaluate all relevant literature, and pool the findings from all high quality studies identified. The goal is to reach a more definitive conclusion than any one study can achieve. Originating in clinical medicine, the systematic review process is exemplified by the work of the Cochrane Collaboration [http://www.cochrane.org/](http://www.cochrane.org/). The process has been adopted in many other fields including public health as illustrated by The Guide to Community Preventive Services [http://www.thecommunityguide.org/](http://www.thecommunityguide.org/). Not to be undertaken lightly, a systematic review can involve evaluating thousands of studies to find that a dozen or so meet the criteria for inclusion.

**University Library System**
The ULS includes Hillman Library and a variety of departmental libraries scattered throughout the Oakland campus. GSPH students often use ULS libraries to obtain information from disciplines such as anthropology, sociology, engineering, biological sciences, and psychology.