Introduction to PubMed

Center for Healthcare Informatics Education

Health Sciences Library
Stony Brook University
8034 SUNY, HSC Level 3 Room 136
Stony Brook, NY 11794-8034
Introduction to PubMed

Effective use of the globally accessible, freely available interface for MEDLINE, the National Library of Medicine’s premier biomedical research, healthcare, and basic medical science bibliographic database

What is PubMed?

PubMed is a World Wide Web database developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM). It was developed in conjunction with publishers of biomedical literature as a search tool for accessing literature citations and linking to full-text journals at Web sites of participating publishers. As a service of the National Library of Medicine, PubMed provides access to over 20 million MEDLINE citations back to the 1950’s and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources. New records are added to PreMedline daily, either directly by publishers or by the National Library of Medicine, making PubMed the most current MEDLINE service available.

What is the difference between MEDLINE and PubMed?

Produced by the National Library of Medicine as well, the MEDLINE database is widely recognized as the premier course for bibliographic and abstract coverage of biomedical literature. MEDLINE encompasses information from Index Medicus, Index to Dental Literature, and International Nursing Index, as well as other courses of coverage in the areas of allied health, biological and physical sciences, humanities and information science as they relate to medicine and healthcare, communication disorders, population biology, and reproductive biology. More than 20 million citations from some 5,400 journals are indexed, plus selected monographs of congresses and symposia. In addition to providing access to MEDLINE, PubMed provides access to:

- The out-of-scope citations (e.g., articles on plate tectonics or astrophysics) from certain MEDLINE journals, primarily general science and chemistry journals, for which the life sciences articles are indexed for MEDLINE.

- Citations that proceed the date that a journal was selected for MEDLINE indexing.

- Some additional life science journals that submit full text to PubMedCentral™ and receive a qualitative review by NLM.

PubMed also provides access and links to the integrated molecular biology databases included in NCBI’s Entrez retrieval system. These databases contain DNA and protein sequences, 3-D protein structure data, population study data sets, and assemblies of complete genomes in an integrated system.
Learning Objectives

This module emphasizes use of PubMed for searching the professional medical journal literature. This is an introductory class that will walk you through the basics of PubMed. Specific learning objectives include:

1. Recognize the nature and interface and be able to effectively use the search tools provided;
2. Navigate the PubMed interface and be able to use the search tools and services provided effectively;
3. Understand Medical Subject Headings (MeSH) and automatic term mapping for controlled vocabulary relating to effective literature searching;
4. Conduct effective search strategies of the biomedical literature content of information resources available through PubMed.

The NCBI Entrez Systems

PubMed is part of the NCBI Entrez System. Entrez is a search and retrieval system that integrates information from databases at NCBI. These databases include nucleotide sequences, protein sequences, macromolecular structures, whole genomes, and MEDLINE.

Entrez provides access to:

- **3D Domains**: domains from Entrez Structure
- **BioSystems**: pathways and systems of interacting molecules
- **Books**: online books
- **Cancer Chromosomes**: cytogenetic databases
- **CDD**: conserved protein domain database
- **dbGaP**: genotype and phenotype
- **EST**: expressed sequence tag records
- **Gene**: gene-centered information
- **Genome**: whole genome sequences
- **Genome Project**: genome project information
- **GENSAT**: gene expression atlas of mouse central nervous system
- **GEO DataSets**: experimental sets of GEO data
- **GEO Profiles**: expression and molecular abundance profiles
- **GSS**: genome survey sequence records
- **HomoloGene**: eukaryotic homology groups
- **Journals**: detailed information about the journals indexed in PubMed and other Entrez databases
- **MeSH**: detailed information about NLM’s controlled vocabulary
- **NLM Catalog**: catalog of books, journals, and audiovisuals in the NLM collections
- **Nucleotide**: core subset of nucleotide sequence records
- **OMIA**: online Mendelian inheritance in animals
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMIM</td>
<td>online Mendelian inheritance in man</td>
</tr>
<tr>
<td>Peptidome</td>
<td>MS/MS proteomic experiments</td>
</tr>
<tr>
<td>PopSet</td>
<td>population study data sets</td>
</tr>
<tr>
<td>Protein</td>
<td>sequence database</td>
</tr>
<tr>
<td>Probe</td>
<td>sequence-specific reagents</td>
</tr>
<tr>
<td>Protein Cluster</td>
<td>a collection of related protein sequences</td>
</tr>
<tr>
<td>PubChem BioAssay</td>
<td>bioactivity screens of chemical substances</td>
</tr>
<tr>
<td>PubChem Compound</td>
<td>unique small molecule chemical structures</td>
</tr>
<tr>
<td>PubChem Substance</td>
<td>deposited chemical substance records</td>
</tr>
<tr>
<td>PubMed</td>
<td>biomedical literature citations and abstracts</td>
</tr>
<tr>
<td>PubMed Central</td>
<td>free, full text journal articles</td>
</tr>
<tr>
<td>Site Search</td>
<td>NCBI web and FTP sites</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SRA</td>
<td>Sequence read archive</td>
</tr>
<tr>
<td>Structure</td>
<td>three-dimensional macromolecular structures</td>
</tr>
<tr>
<td>Taxonomy</td>
<td>organisms in GenBank</td>
</tr>
<tr>
<td>UniGene</td>
<td>gene-oriented clusters of transcript sequences</td>
</tr>
<tr>
<td>UniSTS</td>
<td>markers and mapping data</td>
</tr>
</tbody>
</table>

**Getting Started**

PubMed Services

Journals Database

Use the Journals Database to locate specific indexed journals. Enter the journal's title, MEDLINE abbreviation or ISSN. You may also enter a single word from the journal title or several adjacent words in the title in order to track down the desired journal. Use an asterisk (*) at the end of a word to truncate (e.g., “engl*” would match to “England”, “English”, etc.)

Typing “New Engl**” in the Journals Database is one way to locate the New England Journal of Medicine, for example.
Typing “bone cancer” in the MeSH Database will map the terms to the proper MeSH term (i.e., Bone Neoplasms).

MeSH (Medical Subject Heading) is NLM’s controlled vocabulary for indexing articles in PubMed. MeSH provides a consistent way to retrieve information that may use different terminology for the same concepts. Use the MeSH Database to translate natural language terms to MeSH terms that have been used to index articles in PubMed (see additional information about MeSH below).
The searcher is presented with both the proper MeSH term as well as an indication of where that term resides in the MeSH Tree, which reflects the hierarchal organization of medical subject headings (see below). Once you have located the MeSH term which you wish to conduct your search for citations, click the Send to Search Box with AND button to build the search term into a developing search strategy.

Clicking the Search PubMed button results in a (far too broad) search being conducted for, in this case, all articles that have been indexed with the MeSH term “Bone Neoplasms”.)
Conducting an effective search within PubMed requires executing an effective search strategy. The database is too large – contains far too many citations – for one to use a single search term (even a proper MeSH term) and expect to locate a highly focused article on the desired topic. For example, if we were interested in locating a review article dealing with the chemically induced occurrence of endometrial tissue growth (tumors) as a side effect of Tamoxifen administration in breast cancer therapy, it would not be helpful to simply search on the term “Breast Cancer” (or, to use the proper MeSH term, “Breast Neoplasms”) alone. That said, the database also cannot effectively handle open-ended phrases (e.g., typing “review articles about endometriosis associated with Tamoxifen therapy in breast cancer treatment” will not yield the desired results). To yield the desired results, we need to break down any query into its searchable components (i.e., chemically induced endometriosis; drug therapy for breast cancer; therapeutic use and administration of Tamoxifen). Conduct separate searches for each of these sub-concepts and then use the proper Boolean operator to include only articles that deal with all three concepts. See Principles of Boolean Logic below.

Use the MeSH Database again to narrow your search term more precisely.
Again, click **Send to Search Box with AND** button to execute the narrowed search on drug therapy of breast neoplasms. Then, use the MeSH Database to add additional search terms to the query (i.e., chemically induced endometriosis; therapeutic use and administration of Tamoxifen). Use the MeSH Database to identify the proper MeSH terms for each natural language term (e.g., Breast Neoplasms is the proper MeSH term for breast cancer). For each search term in the search (i.e., Breast Neoplasms, Tamoxifen and Endometriosis), conduct a separate search, then narrow the term in each case according to desired subheadings, and then add the search term to the developing search strategy by clicking on the **Search PubMed** button.

Click the **Advanced Search** and look at the **Search History** to view the results of the developing search strategy.

Use the proper Boolean operators (see below) and syntax to combine the search sets in order to identify articles meeting the desired criteria (i.e., articles indexed about Breast Neoplasms/drug therapy AND Tamoxifen AND Endometriosis/chemically induced).
Click **Preview** to run the strategy to combine the search sets as set forth in the argument (i.e., set #5 AND set #8 AND set #11).

Click the hyperlink entry in the **Result** column to reveal the articles yielded by this search strategy. Or choose to narrow the results further (e.g., narrow the results to review articles).
Note the various parameters by which the searcher may limit the results of a search, including: Type of Article (e.g., review article, meta-analysis, etc), Ages (article subjects), Languages (article language), or Publication Date, etc.

In this case, we are interested only in review articles.

Once the desired parameters have been set, click Go to engage the refined search strategy.

**Note:** the inclusion of each additional parameter will likely result in fewer (and hopefully more relevant) “hits”. In this case, we have narrowed the results of our search down to three (3) review articles for the desired subject matter (narrowed from tens of thousands of articles on the single subject of Breast Neoplasms).
Words to the Wise: Related Citations

Effective literature searching is a product of a carefully designed and executed search strategy, but there is always a degree of iteration involved in refining a search strategy. Serendipity cannot be ruled out! Often a search will turn up a handful of relevant articles among dozens or hundreds of articles otherwise off-topic. By using the Related Citations option, a searcher can leverage one relevant article in a given search by asking PubMed to identify other articles that have been indexed in a similar fashion.

PubMed manages this by searching for articles which have similar Major MeSH headings (see below) as that of the selected article.

---

Effective literature searching is a product of a carefully designed and executed search strategy, but there is always a degree of iteration involved in refining a search strategy. Serendipity cannot be ruled out! Often a search will turn up a handful of relevant articles among dozens or hundreds of articles otherwise off-topic. By using the Related Citations option, a searcher can leverage one relevant article in a given search by asking PubMed to identify other articles that have been indexed in a similar fashion.

PubMed manages this by searching for articles which have similar Major MeSH headings (see below) as that of the selected article.

---

Effective literature searching is a product of a carefully designed and executed search strategy, but there is always a degree of iteration involved in refining a search strategy. Serendipity cannot be ruled out! Often a search will turn up a handful of relevant articles among dozens or hundreds of articles otherwise off-topic. By using the Related Citations option, a searcher can leverage one relevant article in a given search by asking PubMed to identify other articles that have been indexed in a similar fashion.

PubMed manages this by searching for articles which have similar Major MeSH headings (see below) as that of the selected article.
FOCUS and EXPLODE

PubMed presents options to the searcher using MeSH terminology for restricting (focus) to Major MeSH terms only and to determine whether or not to expand the results of a search by including subordinate subject headings falling below a given term in the MeSH Tree (explode). By default, PubMed always explodes a search term.

Exploding a search term will generally result in more articles being included in the search results. The search is broadened conceptually to include not only those articles indexed to a given term (e.g., Bone Neoplasms) but also articles that have been indexed using search terms that are subordinate to a give term in the MeSH Tree (e.g., Adamantinoma, Femoral Neoplasms, Skull Neoplasms, etc).

Restricting (focusing) a search to major topic headings only will generally result in fewer but more relevant search results. PubMed will identify only a subset of those articles which indexers have determined that the given subject heading represents a substantial component of the article’s content.

See below the relative results when options to Restrict Search to Major Topic headings only and to Explode a search term are variously exercised and combined.

Set #18 includes ONLY HIGHLY RELEVANT articles indexed to Bone Neoplasms but NO subordinate concepts that fall beneath Bone Neoplasms in the MeSH Tree (fewer than Set #19).

Set #19 includes ALL articles indexed to Bone Neoplasms but NO subordinate concepts that fall beneath Bone Neoplasms in the MeSH Tree. Set #20 includes ONLY HIGHLY RELEVANT articles indexed to Bone Neoplasms or HIGHLY RELEVANT articles indexed to subordinate concepts that fall beneath Bone Neoplasms in the MeSH Tree (contains more than both Set #18 and Set #19 because additional concepts are being included in the search).

Controlled Vocabulary: “Speaking MeSH”

Effective retrieval of biomedical and healthcare information requires that the searcher understand the principles by which medical information is both organized and accessed. MeSH is a vocabulary of medical and scientific terms assigned to most documents in MEDLINE/PubMed by a team of experts. The advantages of using MeSH terms over text terms found in the title and abstract of a document include:
• **MeSH terms are assigned to the entire article**
  MeSH terms are assigned on the subject of the entire document, not just the citation text (title and abstract). Thus, a search using a MeSH term can find relevant documents even when the term in question is not found in the citation. This is especially important for that minority of documents which do not have abstracts, since titles are very short and often omit important terms.

• **Some MeSH terms are designated as Major Topic**
  Some MeSH terms are assigned as Major Topic terms, meaning that the indexer has determined that these terms represent the major thrust of the document. Using the Major Topic field in PubMed or the MeSH Browser can help you discard documents that are less relevant to your search. These terms are designated in the Citation and MEDLINE reports with an asterisk (*).

• **Some MeSH terms have subheadings**
  Some MeSH terms have an additional subheading, which indicates that a given MeSH term applies in a very specific manner to this document. These subheadings are added to their parent MeSH term with an intervening slash ("/") character. For instance, the term "antibiotics/administration and dosage" means that this document is specifically about the administration and dosage of antibiotics. A search for "antibiotics" will still find this citation, of course. But, judicious use of subheadings can serve to make your query more specific and thus avoid irrelevant material.

• **MeSH terms are Exploded**
  MeSH terms are normally Exploded when used in a search. This means that a search using a MeSH term finds not only documents indexed under that term, but also all documents indexed using more specific terms that are included in the meaning of this MeSH term. For example, searching for the term "Antibiotics" will also retrieve documents indexed using the terms "Antibiotics, Aminoglycoside", "Amoxicillin", "Ticarcillin", and many other terms that fall into this subject. This "Explosion" is accomplished by using the MeSH Tree, which groups specific terms under more general headings. The MeSH Browser permits you to examine the MeSH Tree and select items from it that match your interest area.

  An Important Caveat re: PubMed
  It is important to realize that not all documents in PubMed contain MeSH terms. Assigning MeSH terms takes time; as a result, documents recently added to the database will not have MeSH terms. Moreover, some short documents (e.g. letters to the editor) do not routinely get indexed. Searching using MeSH terms exclusively will not find these documents; thus, if your search fails to turn up enough relevant items, you may wish to substitute appropriate text words in order to find these less well-indexed citations.

**Hierarchical Organization of Medical information: The MeSH Tree**

Once you have selected a term, the MeSH Browser will give you its definition and show you its position in the MeSH Tree, in this fashion:

```
Top of MeSH Tree
  Parent 1
  Parent 2
  Parent 3
  ....
  Term
    Child 1
    Grandchild 1
```

Center for Healthcare Informatics Education  Introduction to PubMed
The Parents shown above are the terms in the MeSH Tree which are each more general in subject than the term below. For instance, "Anti-Infective Agents" is a Parent of "Antibiotics". Similarly, the Children of a term are those terms, if any, that are more specific than the term itself; "Kanamycin" is a Child of "Antibiotics."

If a Child has Children of its own, these will be shown up to a certain point. When space does not permit, a Child will be followed with the character '+' to indicate that it has children which are not displayed.

To move to another term displayed in the MeSH tree, simply select the term.

**Note:** Many terms appear in more than one place in the MeSH Tree, because they logically fall into several different subject areas. When this occurs, the Browser will show you all locations. For example, "Fungemia" is found under both "Septicemia" and "Mycoses" since it matches both subjects.

### Principles of Boolean Logic

<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AND:</strong></td>
<td>search for documents that have both the new term AND the old one</td>
</tr>
<tr>
<td><strong>OR:</strong></td>
<td>search for documents that have either the new term OR the old one</td>
</tr>
<tr>
<td><strong>NOT:</strong></td>
<td>search for documents that have the old term NOT the new one</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AND</strong></td>
<td>HYPERSENSITIVITY and STUDENTS, MEDICAL</td>
</tr>
<tr>
<td></td>
<td>would yield only articles which include both concepts.</td>
</tr>
<tr>
<td><strong>OR</strong></td>
<td>STUDENTS, MEDICAL or EDUCATION, MEDICAL or SCHOOLS, MEDICAL or FACULTY, MEDICAL or ACADEMIC MEDICAL CENTERS</td>
</tr>
</tbody>
</table>

Joins concepts together into one search set and is used to gather synonymous terms or allied concepts together.
Citation Matcher

The citation matcher function allows searchers to match their own list of citations to PubMed citations, using bibliographic information such as journal, volume, issue, page number, and year. The Citation Matcher returns the corresponding PMID. This number can then be used to easily link to PubMed. This service is frequently used by publishers or other database providers who wish to link from bibliographic references on their web sites directly to PubMed citations.

➢ Use the **Single Citation Matcher** to find a single article.

➢ Use the **Batch Citation Matcher** to look for many articles and request them in bulk by e-mail. For instructions, send an empty message except for the word “help” in the body of the message.

Using the partial bibliographic information, we can track down Linus Pauling's publications about Vitamin C and cancer.

**PubMed Single Citation Matcher**

- Use this tool to find PubMed citations. You may omit any field.
- Journal may be the full title or the title abbreviation.
- For first and last author searching, use smith jc format.

- **Journal:**

- **Date:** yyyy/mm/dd (month and day are optional)

- **Volume:**

- **Issue:**

- **Page:**

- **Author name (see help):** Pauling

- **Title words:** Vitamin C and cancer

- **Only as first author**

- **Only as last author**

- **Go**

- **Clear**
Clinical Queries

This specialized search is intended for use by clinicians and has built-in search "filters" based largely upon the work of Haynes, R.B., et al (1994). The filters contain search terms and strategies designed to retrieve particular study types. Five study categories are provided: etiology, diagnosis, therapy, prognosis, and clinical prediction guides. You may indicate whether you wish your search to be a narrow specific search (i.e., include mostly relevant articles but probably omitting a few) or a broad, sensitive search (i.e., include mostly relevant articles but probably including some less relevant ones).

The clinical queries section is highly effective in locating materials in support of evidence-based clinical recommendations. For example, the utility can be used to quickly locate systematic reviews of aspirin therapy for stroke prevention.
If you would like to see details about the search strategies that comprise the filters, go to the Clinical Queries filter table at:

If you are really stressed for time, but still need an evidence-based answer, Clinical Queries should be one of the first places to search. This is, of course, if the question is for etiology, diagnosis, therapy, prognosis, or clinical prediction guides.
LinkOut is a service of Entrez that allows you to link directly from PubMed and other Entrez databases to a wide range of information and services beyond the Entrez system. LinkOut aims to facilitate access to relevant online resources in order to extend, clarify, and supplement information found in the Entrez databases. Examples of LinkOut Resources include full-text publications, biological databases, consumer health information, research tools, and more.

All links are specially assigned to specific database records. When accessing a link through LinkOut, no additional searching should be necessary to access the relevant resource that has been linked to the record. Please encourage online resources that may be valuable to Entrez users to participate in LinkOut.

Click on the applicable tab to access documentation on becoming a LinkOut provider and linking your resources from PubMed and other Entrez databases. Specific documentation is available for libraries, full-text providers and providers of other resources.

Links to external resources are listed in the LinkOut Display of an Entrez record. Users can access the LinkOut display by clicking "LinkOut" from the Abstract and Citation formats in PubMed, or comparable display formats in other Entrez databases. The LinkOut display is also available from the display list of a record. Please consult the Help document for additional information.

In PubMed, external resources can also be accessed through an icon from the Abstract and Citation formats. Users can customize the LinkOut display and icon through the LinkOut Preferences in Cubby.

- LinkOut is a feature that provides external links from PubMed citations to publisher Web sites for full-text journal articles, biological databases, consumer health information, and research tools.

- Books – Provides links from individual PubMed journal citations to full-text of a growing collection of biomedical books.

- Protein – Protein sequences from Swiss-Prot, PIR, PRF, PDB, and translated protein sequences from the DNA sequences databases.

- Nucleotide – DNA sequences from GenBank, RefSeq, and PDB.

- PMC – Provides links from PubMed journal citations to free, full-text of articles in PubMed Central (PMC).

- Structure – The Molecular Modeling Database (MMDB) contains 3-dimensional structures determined by X-ray crystallography and NMR spectroscopy.
• Genome – The genomes represent both completely sequenced organisms and those for which sequencing is in progress.

• Taxonomy – The NCBI taxonomy database contains the names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence.

• OMIM – Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders, with links to literature references, sequence records, maps, and related databases.

My NCBI (formerly known as Cubby)

One advantage of creating an account with MyNCBI, is that you can limit your searches to full text articles available from the Health Sciences Library.
The My NCBI feature provides a Stored Search feature to store and update searches. It also allows you to customize your LinkOut display to include or exclude links to providers.

**To Save Your Searches and Set Up Automatic E-Mail Updates**

1. Sign into My NCBI.

2. Run your search. You may use History numbers in your search, however, PubMed processes from left to right so order your search statement numbers with this in mind, or use parentheses for nesting.

3. Click Save search above the search box.

4. You may edit the name of the search. The maximum number of characters for the search name is 100. Saved search names will be included in the Subject of the automatic e-mail update messages received in the future.

5. Click Save.

6. Fill in the Save Search Settings form if you would like to receive automatic e-mail updates of new search results.

7. Click the set an e-mail address link if you have not already saved one. This address will be used for all e-mail updates for the account.

8. Select the frequency of your updates. E-mailed updates will be sent daily on the scheduled days starting from the appropriate day after the schedule is set.

9. Select a Report format.

10. You may also change the number of items to be sent, or add additional text (up to 200 characters) to the automatic updates.

11. Click Save.

12. To change or create a schedule for a saved search select Saved Data, click the Saved Searches Manage link, and then click Edit for the saved search.

**Note:** Links to Related Citations cannot be saved in My NCBI.

Dates and date ranges are **not** advisable for saved searches. Saved searches cannot be edited.

If an e-mail update bounces back (cannot be delivered) three times, the e-mail address will be changed to invalid.

Each My NCBI account can only have one e-mail address.

In the event an update is not sent out on schedule or is incomplete, the next scheduled update will include the omitted items, if any.
E-mail updates may be affected by e-mail filtering SPAM software. Consider adding nih.gov domain to your e-mail "safe list."

Click the database searches' column headings to sort alphabetically by Name, Last Searched date, or Schedule.