How to develop a search strategy for a Cochrane Review

A. GENERAL INFORMATION

Searches for studies to be included into a Cochrane Review should be extensive and systematic, therefore the search strategy should be a combination of searching text words (in e.g. title, abstract) and of searching keywords that have been indexed by the databases (“controlled vocabulary” e.g. MeSH for MEDLINE; EMTREE for EMBASE).

When using keywords (MeSH, EMTREE) for a search, it is important to keep in mind that indexing in the databases may be incomplete and subjective, thus in certain cases inaccurate. Searching for too specific a category in the database, e.g. “middle-aged” is very unlikely to retrieve all relevant references because the information may be found only in the full text in the majority of the articles and is therefore not indexed. The same applies to fields and subheadings of specific keywords which are organised by the different databases.

Try to be as sensitive as well as specific as possible. Too sensitive a search creates too much workload by resulting in screening unnecessarily many hits. Too specific a search leads to incomplete search results which may bias the systematic review.

The search strategy represents a logic combination of terms for the disease of interest, and the intervention(s). We do not recommend to search for specific outcomes since the increasing complexity of such a search may restrict the results. In Cochrane search strategies, it is advisable to use only the Boolean operators “AND” and “OR”, since “NOT” will also exclude references that contain both an “OR”-search term and a “NOT-term”.

Depending on your question, you may search only for the intervention of interest and leave out the control intervention if you accept all possible control interventions (or “your” control intervention is the most common anyway). Do not search for outcomes since they are usually not well reported in titles or abstracts!


If you do not have access to the newest version of the handbook, please check for necessary corrections under http://www.cochrane.org/resources/handbook/updates4.2.5.htm

Usually, the final search strategies are published in the protocol. The inclusion of search strategies using the appropriate and database-specific search terms and syntax increases the transparency of the review and facilitates future updates.
### B. CHECKLIST FOR DEVELOPING A SEARCH STRATEGY

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Define text words</td>
</tr>
<tr>
<td>2.</td>
<td>Determine synonyms for the text words</td>
</tr>
<tr>
<td>3.</td>
<td>Control for different spellings or using appropriate truncations</td>
</tr>
<tr>
<td>4.</td>
<td>Consider brand names when searching for a specific drug treatment</td>
</tr>
<tr>
<td>5.</td>
<td>Perform test searches – I</td>
</tr>
<tr>
<td>6.</td>
<td>Identify “controlled vocabulary” (keywords) used for the indexing of databases (MeSH for MEDLINE, EMTREE for EMBASE)</td>
</tr>
<tr>
<td>7.</td>
<td>Decide on whether to perform an “exploded” or a “focussed” search for keywords</td>
</tr>
<tr>
<td>8.</td>
<td>Check if all words are spelled correctly!!!!</td>
</tr>
<tr>
<td>9.</td>
<td>Combine logically all search terms</td>
</tr>
<tr>
<td>10.</td>
<td>Perform test searches – II</td>
</tr>
<tr>
<td>11.</td>
<td>Customise the syntax of your search strategy to the specific databases</td>
</tr>
</tbody>
</table>
C. DEVELOPING A SEARCH STRATEGY - A STEP BY STEP APPROACH

1. Define text words
   i. Focus on the disease and the two interventions/procedures you want to compare; do not search for outcomes, age groups of patients etc.
   ii. Make a list of words representing the study population (disease) and the intervention(s)

2. Determine synonyms for the text words
   i. Think of, or search for, synonyms of your search terms
   ii. Use MeSH database for identification of synonyms: check the “entry terms” in the description of a specific MeSH-term

   HOW TO search MeSH-database:
   • Search MeSH with one obvious text word, e.g. chemotherapy
   • Among the suggested MeSH-Terms, choose the one(s) that seem(s) most appropriate to represent your area of research.
     In our review group, this will be rather “antineoplastic combined chemotherapy protocols” instead of “vascunicol”
   • Check the “entry terms” in the description of the MeSH-term whether they represent useful synonyms
     In the example chemotherapy, this would be e.g. antineoplastic agent, anticancer drug

3. Control for different spellings or using appropriate truncations
   i. For the same disease: plasmacytoma or plasmocytoma
   ii. In English: haematological and hematological
   iii. In different languages: leukaemia or leucémie
     This point becomes less relevant since nowadays most of regional journals have at least an English title, but may be important for studies with a long history or dealing with endemic diseases mainly expected to be published in regional medical journals.
   iv. Think of meaningful truncations for e.g. singular/plural; noun/adjective form of search term
   v. Use “wildcards” to replace differences (specific syntax for different databases!)
     e.g. plasm#cytom$ (syntax for MEDLINE via OVID), see below

   EXAMPLE – searching for “erythrocytosis”
   Expected description of your study population:
   - erythrocytosis
   - erythrocytotic patient

   You are not interested in finding studies on
   - erythrocytes
   - erythrocytopenia

   => Use both truncations “erythrocytos*” and “erythrocyt*”
     (syntax: Medline via PubMed)
     ! erythrocyto* would also find erythrocytopenia!
4. **Consider brand names**
   
   
   ii. Think of the possibility of combination drugs (and their brand names) and whether you would like to include them.
   
   iii. Be careful about using abbreviations! Check the results of a test search if there are too many irrelevant hits.

   **EXAMPLE for problematic use of abbreviations**

   ASS = Acetylsalicylacid, Arsenic sulfur (AsS), arthroscopic synovectomy, acute serum sickness, adenylsuccinate synthetase, Equus hemionus (=Asiatic wild ass), American Sociological Society, and many more!!!

5. **Perform test searches - I**
   
   i. Type in text words and / or index terms
   
   ii. Monitor the number of hits after truncations, substitution of wildcards etc. Use correct syntax for each database! E.g. truncation in PubMed by *, in Medline via OVID by $.

   **FIELDS**

   Textwords and keywords are search for in specific fields of the databases, e.g. title, abstract, publication date etc. Different databases may differ with respect to automatically searched fields. The choice of field to be searched influences the specificity of your search. Be careful about routinely searching all fields as e.g. similarities of search terms with author names may lead to very strange results.

   Usually, searching for textwords is limited to the title, original title, abstract and the keyword field which corresponds to *ti, ot, ab, kf* when searching Medline via OVID. For controlled vocabulary in Medline via OVID, “publication type” (pt), or the subject heading (sh) is used.

   *Please refer to the help file of the database to learn about the available fields and their designation.*
6. Identify keywords ("controlled vocabulary" provided by databases)
   i. Map (your list of) text words to the database of controlled vocabulary (e.g. MeSH, EMTREE)
   ii. Study the neighbouring hierarchic levels of your keyword for their usefulness as search terms

HOW TO map your text word to MeSH via PubMed


- Search MeSH database using obvious text word
- Your search term is highlighted in the hierarchical categories of the MeSH indexing system (=MeSH-tree). One term may occur in several trees and may be assigned to different higher categories (=stems) and/or may have different sub-categories (=branches).

Example lymphoma

Stems/Trees:
- "Neoplasms by histologic type"
- "Lymphatic diseases" "/"lymphoproliferative disorders"
- "Immunoproliferative disorders" "/"lymphoproliferative disorders"

Branches:
- See above (or the website) for details (a "+" designates that the term has further branches)
- Pay attention to the different level of comprehensiveness of the different trees!
7. Decide on whether to perform an “exploded” or a “focussed” search

EXPLANATION – “exploded search”

“Explosion” of a term means, that by adding “exp” to your keyword (“exp keyword”) all the lower “branches” of this specific term are automatically included in the search as well. This saves a lot of typing and – if well controlled - can be very useful. Therefore, check both stem and branches of all trees you found with your search term for their usefulness before including (and exploding) index terms. If the branches include too many irrelevant terms, focus your search to your keyword.

In the above mentioned example: “exp lymphoma” will also search for references which are indexed as “histiocytosis, malignant” – do you want to be as comprehensive?

CAVE
Default settings and syntax between different databases may be different!
e.g.: searching MEDLINE using MeSH terms in PubMed per default explodes the term, not so in OVID!

NOTE: If your search platform is PubMed, the mapping to the keywords is performed automatically for all terms which are entered without further specification! See http://www.ncbi.nlm.nih.gov/books/bv.fcgi?id=helppubmed.section.pubmedhelp.Appendices#pubmedhelp.How_PubMed_works_aut
In OVID, you have to write out both textwords and keywords in order to search for them

i. Are there other terms in the stems or branches that should be included (e.g. as text words?)
ii. Instead of searching all parallel branches of one category, add “exp” to the upper term (closer to the stem) – saves typing!
iii. Are there terms in the lower branches that will definitely increase the noise in your search? Weigh the impact in terms of references to screen. Either do not explode the upper term and use individual parallel branch as search terms, or explode nevertheless if the “noise” will not have a big impact (e.g. a very rare condition).
iv. Are there terms in the tree that may be used interchangeably by indexers? There is some subjectivity in the indexer whether she/he will use THROMBOCYTE AGGREGATION or THROMBOCYTE AGGREGATION INHIBITION if the publication is on antiplatelet drugs. Include both keywords in your strategy.

8. Spell check ALL search terms!!!

9. Combine logically all search terms
i. Structure your search terms: group both text words and keywords for the disease, the intervention(s) and the control, respectively.
ii. Combine all relevant text words and keywords of the disease of interest by “OR”
iii. Combine all relevant text words and keywords of your intervention(s) by “OR”
iv. (Combine all relevant text words and keywords of your control(s) by “OR”)
v. Combine “disease” AND “intervention” (AND “control”)
vi. Add Cochrane RCT filter: “disease” AND “intervention” (AND “control”) AND RCT
vii. (Limit your search to publication dates for recently licensed drugs.)
10. Perform test searches - II
   i. Type in your logically combined search terms first individually
   ii. Monitor the number of hits after logical combination of the search terms
   iii. NOTE: The combination with the Cochrane-RCT-filter usually decreases drastically the number of references!

11. Customise the syntax of your search strategy to the specific databases
Different databases (e.g. MEDLINE, EMBASE) or different database providers (e.g. OVID) are very different concerning the usage of wildcards, possibilities of truncations, automatic mapping functions to controlled vocabulary, punctuation etc. and a search strategy developed for MEDLINE via OVID cannot be used for searching MEDLINE via PubMed without “translation”. Check the help files of the databases to look for instructions.

The editorial base of the CHMG usually develops search strategies for MEDLINE using OVID. In order to be able of offer you optimal assistance we would advise you to use OVID as well. Ask your librarian whether your institution has a subscription to this database provider.

   i. A Medline-search will not find all relevant references in EMBASE and vice versa.
   ii. A search developed for MEDLINE via OVID does not work in PubMed without “rephrasing and reformatting” (see below)
   iii. CAVE: MeSH-Terms are not identical to EMTREE terms! Mapping of keywords to EMTREE requires an access to the commercial EMBASE database.
   iv. Some databases defined “stopwords” – terms which cannot be searched, e.g. “ALL” for acute lymphoblastic leukaemia, cannot be searched in PubMed. It is interpreted as “everything” instead of a specific type of leukaemia.
   v. Hyphens are interpreted differently between databases. Sometimes they are ignored, sometimes not, which necessitates searching for both versions of a hyphenated word (e.g. Non-hodgkin or Non?hodgkin or Nonhodgkin)

**EXPLANATION and EXAMPLE– Syntax**

All databases (e.g. MEDLINE, EMBASE) or database providers (e.g. OVID, Silverplatter) have a specific syntax for searching which is different with respect to wildcards, indexing and “punctuation”.

Searching for plasmocytoma:

OVID: # - substitution of a single character by another, $ - truncation
“Plasm##ytom$” finds: plasmacytoma, plasmocytoma, Plasmazytom, Plasmozytom

PubMed: no substitution within words possible, therefore: plasmacytom* or plasmocytoma, Plasmazytom, Plasmozytom
D. EXAMPLES OF SEARCH STRATEGIES ALREADY AVAILABLE

i. **CHMG search Medline via OVID**

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1950 to Present

<table>
<thead>
<tr>
<th>#</th>
<th>Search History</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HEMATOLOGIC DISEASES/</td>
<td>10090</td>
</tr>
<tr>
<td>2</td>
<td>exp HEMATOLOGIC NEOPLASMS/</td>
<td>4613</td>
</tr>
<tr>
<td>3</td>
<td>(h?ematolog$ adj1 malignan$).tw,kf,ot.</td>
<td>261982</td>
</tr>
<tr>
<td>4</td>
<td>(h?ematolog$ adj1 neoplas$).tw,kf,ot.</td>
<td>184946</td>
</tr>
<tr>
<td>5</td>
<td>exp BONE MARROW DISEASES/</td>
<td>59574</td>
</tr>
<tr>
<td>6</td>
<td>exp LYMPHOMA/</td>
<td>141272</td>
</tr>
<tr>
<td>7</td>
<td>exp LEUKEMIA/</td>
<td>153521</td>
</tr>
<tr>
<td>8</td>
<td>hodgkin$.tw,kf,ot.</td>
<td>40570</td>
</tr>
<tr>
<td>9</td>
<td>lymphogranulomatous.tw,kf,ot.</td>
<td>1832</td>
</tr>
<tr>
<td>10</td>
<td>lymphom$tw,kf,ot.</td>
<td>88125</td>
</tr>
<tr>
<td>11</td>
<td>histiocyt$tw,kf,ot.</td>
<td>16654</td>
</tr>
<tr>
<td>12</td>
<td>granulom$tw,kf,ot.</td>
<td>39215</td>
</tr>
<tr>
<td>13</td>
<td>non-hodgkin$tw,kf,ot.</td>
<td>21704</td>
</tr>
<tr>
<td>14</td>
<td>nonhodgkin$tw,kf,ot.</td>
<td>83</td>
</tr>
<tr>
<td>15</td>
<td>reticulosis.tw,kf,ot.</td>
<td>1298</td>
</tr>
<tr>
<td>16</td>
<td>reticulosarcom$tw,kf,ot.</td>
<td>987</td>
</tr>
<tr>
<td>17</td>
<td>(burkitt$ adj (lymphom$ or tumo?r$)).tw,kf,ot.</td>
<td>5610</td>
</tr>
<tr>
<td>18</td>
<td>lymphosarcom$tw,kf,ot.</td>
<td>3810</td>
</tr>
<tr>
<td>19</td>
<td>brill-symmer$tw,kf,ot.</td>
<td>200</td>
</tr>
<tr>
<td>20</td>
<td>plasm##ytom$tw,kf,ot.</td>
<td>5382</td>
</tr>
<tr>
<td>21</td>
<td>myelom$tw,kf,ot.</td>
<td>32697</td>
</tr>
<tr>
<td>22</td>
<td>sezary.tw,kf,ot.</td>
<td>1299</td>
</tr>
<tr>
<td>23</td>
<td>leuk?em$tw,kf,ot.</td>
<td>154349</td>
</tr>
<tr>
<td>24</td>
<td>myelodysplas$tw,kf,ot.</td>
<td>8966</td>
</tr>
<tr>
<td>25</td>
<td>aplast$ an?em$tw,kf,ot.</td>
<td>6292</td>
</tr>
<tr>
<td>26</td>
<td>or/1-25</td>
<td>762092</td>
</tr>
<tr>
<td>27</td>
<td>randomized controlled trial.pt.</td>
<td>227471</td>
</tr>
<tr>
<td>28</td>
<td>controlled clinical trial.pt.</td>
<td>73813</td>
</tr>
<tr>
<td>29</td>
<td>randomized controlled trials/</td>
<td>45965</td>
</tr>
<tr>
<td>30</td>
<td>random allocation/</td>
<td>56536</td>
</tr>
<tr>
<td>31</td>
<td>double blind method/</td>
<td>88657</td>
</tr>
<tr>
<td>32</td>
<td>single blind method/</td>
<td>10442</td>
</tr>
<tr>
<td>33</td>
<td>or/27-32</td>
<td>385819</td>
</tr>
<tr>
<td>34</td>
<td>(ANIMALS not HUMANS).sh.</td>
<td>3010245</td>
</tr>
<tr>
<td>35</td>
<td>33 not 34</td>
<td>362976</td>
</tr>
<tr>
<td>36</td>
<td>clinical trial.pt.</td>
<td>431076</td>
</tr>
</tbody>
</table>
How to develop a search strategy

37 exp clinical trials/ 184553
38 (clin$ adj25 trial$).ti,ab. 128305
39 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab. 89571
40 placebos/ 25632
41 placebo$.ti,ab. 100908
42 random$.ti,ab. 371144
43 research design/ 45464
44 or/36-43 835375
45 44 not 34 777514
46 45 not 35 430781
47 35 or 46 793757
48 26 and 47 40551
49 limit 48 to ed=20060101-20061231 2808

#1-25: searching for haematological malignancies
#26-47: Cochrane RCT filter (Higgins 2006)
#49: limitation to a specific entry date into the database (year 2006 only)

ii. Autologous stem cell transplantation in acute lymphoblastic leukaemia

Search Strategy for MEDLINE via OVID

1 exp LEUKEMIA, lymphocytic, Acute/
2 leuk?emi$.ab,ot,ti,kw.
3 (akut$ or acut$).ab,ot,ti,kw.
4 2 and 3
5 1 or 4
6 BONE MARROW TRANSPLANTATION/
7 HEMATOPOIETIC STEM CELL TRANSPLANTATION/
8 PERIPHERAL BLOOD STEM CELL TRANSPLANTATION/
9 transplant$.tw,kf,ot.
10 graff$.tw,kf,ot.
11 or/6-10
12 autolog$.tw,kf,ot.
13 homolog$.tw,kf,ot.
14 12 or 13
15 11 and 14
16 exp TRANSPLANTATION, AUTOLOGOUS/
17 exp TRANSPLANTATION, HOMOLOGOUS/
18 (autotransplant$ or auto-transplant$).tw,kf,ot.
19 (autograft$ or auto-graft$).tw,kf,ot.
20 (homograft$ or homo-graft$).tw,kf,ot.
21 (homotransplant$ or homo-transplant$).tw,kf,ot.
22 or/15-21
23 5 and 22
+ RCT filter tested by the Cochrane Collaboration Higgins 2006

Search strategy for EMBASE via OVID

1 exp ACUTE LYMPHOBLASTIC LEUKEMIA/
2 exp ACUTE LYMPHOCYTIC LEUKEMIA/
3 leuk?emi$.ab,ot,ti.
4 (akut$ or acut$).ab,ot,ti.
5 3 and 4
6 1 or 2 or 5
7 exp STEM CELL TRANSPLANTATION/
8 transplant$.ab,ot,ti
How to develop a search strategy

Cochrane Haematological Malignancies Group - Naumann 2007

9 graft$,.ab,ot,ti.
10 Or/7-9
11 autolog$,.ab,ot,ti.
12 homolog$,.ab,ot,ti.
13 11 or 12
14 10 and 12
15 AUTOTRANSPLANTATION/
16 AUTOGRaFT/
17 (autograft$ or autotransplant$).ab,ot,ti.
18 (homograft$ or homotransplant$).ab,ot,ti
19 Or/14-18
20 6 and 19
21 Randomized Controlled Trial/
22 Randomization/
23 Double Blind Procedure/
24 Single Blind Procedure/
25 Or/21-24
26 Animal/ not Human/
27 25 not 26
28 Clinical Trial/
29 (clin$ adj25 trial$).ti,ab.
30 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.
31 Placebo/
32 placebo$.ti,ab.
33 random$.ti,ab.
34 Or/28-33
35 34 not 26
36 35 not 27
37 27 or 36
38 20 and 37

Search strategy for CENTRAL

1 MeSH descriptor Leukemia, Lymphocytic, Acute explode all trees in MeSH products
2 leukemi* OR leukaemi* in Record Title or leukemi* OR leukaemi* in Abstract or leukemi* OR leukaemi* in Keywords
3 akut* OR acut* in Record Title or akut* OR acut* in Abstract or akut* OR acut* in Keywords
4 (#2 and #3)
5 (#1 or #4)
6 MeSH descriptor Stem Cell Transplantation explode all trees in MeSH products
7 transplant* in Record Title or transplant* in Abstract or transplant* in Keywords
8 graft* in Record Title or graft* in Abstract or graft* in Keywords
9 (#6 or #7 or #8)
10 Autolog* in Record Title or autolog* in Abstract or autolog* in Keywords
11 homolog* in Record Title or homolog* in Abstract or homolog* in Keywords
12 (#10 or #11)
13 (#9 and #12)
14 MeSH descriptor Transplantation, Autologous, this term only in MeSH products
15 MeSH descriptor Transplantation, Homologous, this term only in MeSH products
16 autograft* OR autotransplant* in Record Title or autograft* OR autotransplant* in Abstract or autograft* OR autotransplant* in Keywords
17 homograft* OR homotransplant* in Record Title or homograft* OR homotransplant* in Abstract or homograft* OR homotransplant* in Keywords
18 (#13 or #14 or #15 or #16 or #17)
19 (#5 and #18)