

## Is it possible to focus EMTREE without loss of sensitivity when searching Embase for systematic reviews? Analysis of Cochrane Reviews and HTA reports

Steven Duffy<sup>1</sup>, Janine Ross<sup>1</sup>, Kate Misso<sup>1</sup>, Shelley de Kock<sup>1</sup>, Caro Noake<sup>1</sup> and Lisa Stirk<sup>1</sup>

<sup>1</sup>*Kleijnen Systematic Reviews, York, United Kingdom*

Corresponding author: Steven Duffy, [steven@systematic-reviews.com](mailto:steven@systematic-reviews.com)

### Abstract

**Introduction:** Systematic reviews and health technology assessments require a comprehensive search of numerous databases in order to minimise bias. MEDLINE and Embase are the most commonly searched bibliographic databases when undertaking systematic reviews of health care interventions. As the overall search results for systematic reviews appear to be getting increasingly larger, it would reduce workload and expedite completion if search results could be made smaller and more relevant. Focusing literature searches to ‘Major’ EMTREE subject heading terms in Embase could significantly reduce the number of records retrieved.

**Objectives:** To investigate whether restricting EMTREE indexing terms to focus when searching Embase reduces the number of records retrieved without loss of relevant studies.

**Methods:** Embase searches conducted in recent Cochrane Reviews and UK National Institute for Health Research (NIHR) Health Technology Assessment reports were retrospectively compared with search strategies in which the EMTREE terms had been focused. The records retrieved by the focused EMTREE search were investigated to see if included studies identified by the original unrestricted Embase search strategy were still retrieved.

**Results:** The data collected were analysed to identify: total with and without restriction to focus, yield of included records, and Number Needed to Read (NNR) to detect relevant references.

**Conclusions:** The investigation explored overall yield and recall of relevant included records by each approach, and whether focussing EMTREE terms reduces screening burden without significantly impairing recall of relevant records. Reducing the number of records retrieved from systematic review searches without a loss of sensitivity would improve efficiency, save time and minimise costs. However, the results of this investigation were inconclusive, other than to suggest that restricting EMTREE to focus should only be used with caution.

*Key words: Abstracting and Indexing as Topic; Databases, Bibliographic; Information Storage and Retrieval; Review Literature as Topic; Technology Assessment, Biomedical; Vocabulary, Controlled*

### Introduction

Systematic reviews and health technology assessments require comprehensive searches of multiple databases in order to minimise bias. MEDLINE and Embase are the most commonly searched bibliographic databases when undertaking systematic reviews of health care interventions. As the overall search results for systematic reviews appear to be getting increasingly larger, it would reduce workload and expedite completion if search results could be made smaller and more relevant. We wanted to see if focusing literature searches to ‘Major’ EMTREE subject heading terms in Embase could significantly reduce the number of records retrieved, without losing potentially relevant studies.

Where MEDLINE averages between 10 and 15 MeSH (Medical Subject Headings) index terms with each record, it is not unusual for Embase to include up to 50 (and often more) Emtree subject heading index terms with each record.[1] This exhaustive indexing of records in Embase can lead to the retrieval of large numbers of irrelevant records. However, indexing terms from the Emtree thesaurus can be restricted to retrieve results where the subject heading term is the main focus of the article (Restricting to Focus (RTF)). Embase records average around 3-4 of these focused 'Major' Emtree terms. Elsevier, the producers of Embase, suggest that focusing Emtree subject heading terms in Embase significantly reduces the number of records retrieved by limiting retrieval to the most relevant records.[1]

Previous investigations undertaken by Information Specialists at the German Institute for Quality and Efficiency in Health Care (IQWiG) indicated that focusing Emtree in searches specifically for reviews of drug interventions retained sensitivity while dramatically reducing the number of records retrieved.[2] A more recent study suggests that search strategies used for topics other than drug interventions should carry out sensitivity tests before using RTF, whilst also recommending that caution should be used if looking to apply RTF within the population concept, and if planning to use RTF with more than one search concept.[3]

### Objectives

Our initial investigations involved retrospectively comparing Embase searches for six reviews (and a review update) undertaken at Kleijnen Systematic Reviews (KSR) with search strategies in which the Emtree terms were RTF. The records retrieved by the focused Emtree searches were investigated to see if included studies identified by the original unfocused Embase searches were still identified. In all cases the number of records retrieved with focused Emtree searches was reduced, in some cases significantly, and in only two cases with the loss of any included studies.

Unfortunately, the results of this earlier investigation were inconclusive. The topics covered were diverse with little similarity in review question or search terms used, and so the results were not generalizable to all systematic literature searches. We wanted to investigate whether focusing Emtree across a larger set of comparable reviews produced more conclusive results, so that Information Specialists can confidently use focused Emtree in their search strategies. We decided to investigate what are recognised as the *gold standard* in systematic reviewing, Cochrane Reviews, as well as the long established Health Technology Assessment (HTA) Reports produced by the UK National Institute for Health Research (NIHR). Once again, the Embase searches were retrospectively compared with search strategies in which the Emtree terms had been focused. The records retrieved by the focused Emtree search were investigated to see if included studies identified by the original unrestricted Embase search strategy were still identified.

### Methods

We searched for Cochrane reviews in the Cochrane Database of Systematic Reviews (CDSR) using the term *Embase* and the limit: *Publication Year from 2010 to 2015, in Cochrane Reviews (Reviews only)*. This search retrieved 3629 Cochrane reviews, from which we randomly selected 50 using <http://www.randomizer.org/>.

We searched the Health Technology Assessment (HTA) database in the Cochrane Library for NIHR HTA reports, but this approach did not work. Unfortunately, the majority of records retrieved could not be used for this investigation as they were not based on systematic reviews. Instead, the results included details of ongoing projects, reports of trials, National Horizon Scanning Centre reports, and other unusable citations. A second approach was taken which used the following search strategy in PubMed to ensure NIHR HTA reports of systematic reviews were identified:

- #1 "Health Technol Assess"[jour]
- #2 ("2010"[Date - Publication] : "3000"[Date - Publication])
- #3 systematic\*[ti]
- #4 (#1 AND #2 AND #3)

As with CDSR, 50 records were randomly selected (using <http://www.randomizer.org/>). The two sets of 50 randomly selected reviews were then screened using the following inclusion criteria:

- Embase included in the literature searches
- Embase searched via Ovid
- Date restriction (2010-2015)
- Did not 'restrict to focus' Emtree
- Did not only search the Cochrane Group Trials Register
- Free text used ti,ab or tw NOT sh, hw, mp or af
- Search strategy no longer than 70 lines
- No more than 20 included studies

## Results

Only 16 of the 50 randomly selected Cochrane reviews met the inclusion criteria. Reviews were excluded because: Embase host was not Ovid (11 reviews); Cochrane Group trials register was searched (10 reviews); the field tag *mp* was used in the search strategy (8 reviews); the Embase search strategy was not reported (5 reviews).

Details of the search yield from both the original search strategy compared to the RTF search strategy are presented in Table 1. Overall we found that:

Search yield: 40% average fewer records retrieved with RTF

Sensitivity: original searches 95.5%; RTF searches 94%

Number needed to read: original searches 308; RTF searches 195

Table 1. Cochrane reviews: search yield

Cochrane Review	Included Studies: Total	Included Studies: Embase	Original search results	RTF search results	Difference in search yield
Akhtar	3	3	727/3	183/3	544 (74%)
Akl	15	14	4730/13	2413/13	2317 (49%)
Calladine	16	16	1120/14	909/11	211 (19%)
Cheong	20	17	559/17	308/17	251 (45%)
Cruciani	5	4	128/4	53/4	75 (59%)
Dwan	16	16	4275/10	3620/10	655 (15%)
FlorCruz	12	10	386/10	195/10	191 (49%)
Giljaca	18	14	20020/14	11971/14	8049 (40%)
Hosking	17	5	3260/5	3173/5	87 (3%)
McGee	7	6	257/6	95/6	162 (63%)
Rueda	15	15	1067/14	639/12	428 (40%)
Shang	5	5	244/5	85/5	159 (65%)
Vasudex	6	4	168/4	151/4	17 (10%)
Veltman	14	13	969/12	860/12	109 (11%)
Wakai	4	4	1787/4	709/4	1078 (60%)
Zhang	6	4	3573/4	2032/4	1541 (43%)

Just 17 of the 50 HTA reports met the inclusion criteria. Searches were excluded because: the Embase search strategy was not reported (17 HTA reports); Embase host was not Ovid (2 HTA reports); the field tag *mp* or *af* was used in the search strategy (8 HTA reports); the field tag *sh* was used in the strategy (1 HTA report); EMTREE terms were not available (1 HTA report); did not search Embase (2 HTA reports); EMTREE terms were already RTF (2 HTA reports).

For details of the HTA reports search yield see Table 2. Overall, our investigation found that:

Search yield: 37% average fewer records retrieved with RTF

Sensitivity: original searches 87%; RTF searches 79%

Number needed to read: original searches 398; RTF searches 260

Table 2. HTA reports: search yield

HTA Report	Included Studies: Total	Included Studies: Embase	Original search results	RTF search results	Difference in search yield
Brown	24	24	3680/24	3483/24	197 (5%)
Edwards	11	9	1831/9	1500/9	331 (18%)
Frampton	28	27	6084/23	4047/23	2037 (33.5%)
Gibbons	22	22	478/16	74/7	404 (85%)
	15	15	479/8	49/5	430 (90%)
	16	16	7020/14	1035/8	5985 (85%)
	10	10	1112/6	26/0	1086 (98%)
Greenhalgh	1	1	2964/1	2109/1	855 (29%)
Hislop	13	12	274/12	229/12	45 (16%)
Main	23	23	17496/19	10186/17	7310 (42%)
Maund	31	28	8499/28	5123/28	3376 (40%)
McKenna	10	9	1895/8	835/8	1060 (56%)
Meadows	20	18	2793/17	2229/17	564 (20%)
Mowatt	31	28	4038/27	2701/26	1337 (33%)
Mowatt	65	58	6120/56	5413/56	707 (12%)
Owen	19	16	5799/8	4439/8	1360 (23%)
Simmonds	23	21	5278/15	4416/15	862 (16%)
	34	31	3064/28	2630/28	434 (14%)
Waugh	9	9	4041/9	2826/9	1215 (30%)
Westwood	22	17	1316/17	1106/17	210 (16%)
Westwood	33	30	11245/30	8827/29	2418 (22%)

## Conclusions

Intriguingly, our main findings revolved around issues that had not been anticipated beforehand. We specifically chose Cochrane reviews and NIHR HTA reports to investigate as these are recognised as the *gold standard* when it comes to systematic reviews and health technology assessments, and anticipated comparing high quality search strategies. We found that the search strategies were suboptimal, with many weaknesses. In numerous cases there were no Embase searches at all, or if Embase was searched, the search strategy was not reported. It was not unusual to see multi-file database searching where MEDLINE and Embase had been searched simultaneously. Individual Cochrane Group Trials Registers were searched in preference to specific bibliographic databases. Then, the search strategies themselves often included mistakes in syntax and set combinations, did not use truncation or proximity operators sufficiently, lacked synonyms, and were limited by date or language. There were actually more questions raised about the quality of systematic review searching during this investigation than answers about the potential use of RTF EMTREE.

The search strategies, especially those used by the Cochrane reviews, retrieved relatively small numbers of records, and it was not surprising to see that they had identified all (or most) of the subsequently included studies. It would be interesting to see if additional included studies would be identified with better designed search strategies.

We planned further, larger samples, but similar investigations had been conducted concurrently and subsequently published by the Canadian Agency for Drugs and Technologies in Health (CADTH) [3]. The CADTH report suggests that information specialists should use caution when considering RTF Emtree, especially when focusing the population concept, or focusing more than one concept. Searchers should be particularly confident of the high sensitivity of their search strategy.

From our investigations we would agree with the findings of the CADTH report. Information specialists should be confident of the quality and sensitivity of their search strategy before even considering RTF Emtree. We suggest that RTF Emtree should only be used once all means of reducing an extremely large number of records retrieved (unmanageable in the context of time and resources available) have been exhausted. Further, information specialists should ensure that they compensate for using RTF by undertaking more sensitive searching elsewhere, either with the search strategy design itself, or through searches in other databases and resources.

#### **Footnote: Field tags**

*hw* includes the ‘Heading Word’ field option where a single word is searched for in Emtree. This negates focussing Emtree as it does not differentiate between the two options, e.g. (endometri\$ adj3 tumor\$).mp will search for the Emtree term endometrium tumor/ whether it is RTF or not.

*af* includes ‘all fields’

*sh* includes the ‘Subject Heading’ field, containing Emtree terms

*tw* textword field is an alias for all fields in a database which contain text, and includes title and abstract in Embase.

*mp* enables ‘multi-purpose’ searching without specifying a particular field, in Ovid Embase mp includes: title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword.

#### **REFERENCES**

1. Elsevier. Embase indexing guide 2015: a comprehensive guide to Embase indexing policy [Internet]: Elsevier, 2015 [accessed 12.2.16] Available from: [https://www.elsevier.com/\\_\\_data/assets/pdf\\_file/0016/92104/Embase-indexing-guide-2015.pdf](https://www.elsevier.com/__data/assets/pdf_file/0016/92104/Embase-indexing-guide-2015.pdf)
2. Hausner E, Simon M, Kaiser T. Focused Emtree terms vs. non-focused Emtree terms to search for clinical drug trials in Embase. Paper presented at 16th Cochrane Colloquium: evidence in the era of globalisation; 3-7 Oct 2008; Freiburg: Germany. 2008.
3. Glanville J, Kaunelis D, Mensinkai S, Picheca L. Pruning Emtree: does focusing Embase subject headings impact search strategy precision and sensitivity? [Internet]. Ottawa: CADTH, 2015 Apr. [cited 2016-04-27]. Available from: <https://www.cadth.ca/pruning-mtree-embase>