NICE Evidence search: notes for advanced searchers

NICE Evidence search is a search engine that works in a similar way to search engines such as Google, rather than like a bibliographic database. The approach taken to searching within it needs to be different to the approach taken for a traditional literature search. Evidence search uses the full text of documents within its index to produce ranked sets of results, rather than relying wholly on structured metadata. It is also constantly updated, rather than a fixed database of information, so searchers may find different sets of results each time they search.

This document provides some insight to how Evidence search works, with an explanation of what happens when search features are used. It covers ranking of search results, word stemming (lemmatisation), synonym expansion and wildcard searching, phrase searching, Boolean operators, stop words, spelling correction, and an explanation of why there is no UK versus international option for limiting results.

Ranking of results – relevance

Evidence search brings results back in a ranked order, aiming to bring the most relevant evidence to the top of the search results page. Three different elements are used in the ranking algorithm:

- Where search terms appear in a particular document:
  - position: whether or not they appear in the title and whether they come near the beginning or end of the document
  - frequency: how often they appear in the full text
  - proximity: how close the terms are to each other.

- What type of information a particular document is:
  - guidance documents are given a ranking ‘boost’, so that they come nearer the top of the search results
  - the hierarchy of evidence is then followed, so that systematic reviews and other summaries of evidence come higher up the results than primary research.

- How up to date a particular document is:
  - generally, more recent publications appear higher up the list of search results.

All three of these elements are combined, so that the overall ranking of a document is decided by a combination of the occurrences of the search terms, the information type and when it was published.
Lemmatisation

Lemmatisation is an advanced form of word stemming to ensure documents that contain different forms of the search terms are included in results. In particular, it ensures that if a search on the singular form of a word is carried out, the results will also contain documents with the plural form, even when these are irregular.

Lemmatisation means that different variations of a word do not need to be added as separate search terms, for example:

- typing **foot** in the search box will return documents that contain either **foot** or **feet**
- typing **pregnancy** in the search box will return documents that contain either **pregnancy** or **pregnancies**.

Synonym expansion

Evidence search automatically expands synonyms, including common abbreviations and American spellings. For example:

- if **COPD** is entered as a search term, the results will also include documents that contain the phrase **chronic obstructive pulmonary disease**
- if **fetal** is entered, documents for **foetal** will also be returned.

The list of synonyms and abbreviations for Evidence search is substantial, but it can be added to, so please do send additional suggestions.

Wildcard searching

The lemmatisation rules used in Evidence search mean that searchers will not often need to use the wildcard character (*) to search for word stem variations.

The wildcard character can be used on a search term, but it should be used with caution because it affects the ranking of search results. If the character is used to find variations of a word, the search term that the wildcard is applied to is ignored in the ranking, because Evidence search cannot determine which variation of the word to rank on. This might mean that documents that would have been expected to appear high up on a list of search results will appear lower down. Searchers would need to be prepared to spend time looking through all of the search results if the wildcard character is used.
Phrase searching

Evidence search supports the use of quotation marks to allow phrase searching. Quotation marks should be used with care because using them ‘switches off’ search features such as lemmatisation, spellcheck and synonym expansion.

A good case for using phrase searching would be when looking for a specific known item (perhaps a title) such as “MI secondary prevention”, the title of a CKS topic.

However for ordinary phrases, the ranking policies will tend to ensure that documents where search terms occur close to each other naturally come near the top of the results. For example, if a search for diabetes type 2 is carried out, results for diabetes type 2 and type 2 diabetes are likely to appear near the top of the results. If the phrase “diabetes type 2” is entered, the results would miss documents containing the phrase type 2 diabetes.

Boolean operators

In cases where more than one search term is added to the search box, Evidence search automatically combines search terms using AND.

An OR search can be carried out by typing the word OR between the search terms; for example: elderly or older (please note, OR does not have to be in capital letters).

Brackets can be used to create and combine sets of search terms such as (elderly or older) and (exercise or fitness).

Evidence search does not support the use of the Boolean operator NOT. This is because the NICE guideline development process often produces “do not do” recommendations. If Evidence search used the NOT operator, then documents containing the phrase “do not do” would not be findable.

Stop words

Evidence search does not look for occurrences of certain ‘stop’ words, even if they are included in the search (unless they are used as part of a phrase in quotation marks). This is a list of the stop words used in Evidence search:

- a
- an
- any
- are
- as
- at
- be
- but
- by
- can
- etc
- for
- had
- has
- have
- here
- I
- if
- in
- into
- is
- it
- its
- nor
- of
- on
- our
- so
- than
- that
- the
- then
- there
- they
- this
- those
- to
- we
- were
- what
- when
- were
- was
Spelling corrections

Evidence search automatically corrects spelling errors. If it cannot match a word used in a search against an entry in the spellcheck dictionary, the word is corrected to the best match available and documents are returned for that word instead. For example, if diabetes is typed into the search box, Evidence search will search instead for diabetes.

Evidence search always gives searchers the option to reverse its spelling corrections. For example, if someone wanted to search for daibetes, they can use a link above the results to find matches with the spelling correction turned off:

UK versus international search

It is not possible in Evidence search to limit search results to just UK or international documents, for two reasons:

- Many of the websites that are crawled for Evidence search do not provide geographical metadata on a document-by-document basis. This means that all documents from an evidence source would have to be assigned to either ‘UK’ or ‘international’ categories, which would often be inaccurate. On the other hand, if Evidence search were to use only explicitly declared geographical metadata, the categorisation would be incomplete. If geographical area was then used to filter, uncategorised documents would be missing from the search results.

- Definitions of ‘UK’ and ‘international’ are hard to agree on. Should the definitions be based on the origin of a document or its intended audience? The geographical origin of documents on the web may be hard to define (the author, the funder, and the publisher or host of a document may all come from different countries). And, in the case of guidance in particular, a document may be suitable and applicable to both UK and international audiences. Guidance produced by the European Society of Cardiology, for example, is applicable to England and Wales and to other European countries.

If you have any queries about the search features of NICE Evidence search, or would like to provide any feedback, please email nice@nice.org.uk.