

## [Around meta-analysis \(13\): Benchmarking your online database searches](#)

30/9/2021

By Malgorzata Lagisz

If you are conducting systematic literature searches you are very likely to run searches in online databases. How well these searches capture relevant literature?

In our recent paper ([Foo et al. 2021](#)) we explained the systematic review process, from question formulation to screening the retrieved records, but we only briefly mentioned the need for checking the performance of database searches. There is very little practical guidance available on how to do this and there is also some confusion about the relevant terminology.

To fill this gap, I will explain what benchmarking is and how it can be done using Scopus database search as an example. There are some more complicated ways of benchmarking, esp. in computational projects, but I leave them for another time, as usually simple benchmarking will be sufficient (unless you get a very picky manuscript reviewer).

### **Why Scopus?**

Scopus has a broad and interdisciplinary coverage, user-friendly interface with advanced search functionality. For these reasons, Scopus is our database of choice for performing main searches. We also use Web of Science (Web of Knowledge), which is quite similar, and techniques presented in this blog will apply there too, with minor modifications.

### **What is benchmarking?**

Its main function is to check if your search is able to find relevant papers indexed (i.e. available as records) in a given database. You do not know in advance all the relevant records in the database until you check ALL of the records there against your inclusion criteria (and this is obviously impossible). So, checking is usually done against a pre-determined set of relevant publications. We call this set "benchmarks" or "benchmarking set", but other names used for these are "validation set" or "test set".

### **How to get a benchmarking set of papers?**

We would usually collect relevant papers from various sources - relevant reviews on the topic, [Google Scholar](#), [ConnectedPapers](#), [CiteHero](#), etc. We pick a set of minimum 10 papers, and collect their full references (bibliographic records), including DOI.

### **How to quickly benchmark in Scopus?**

I use an example based on our recent protocol of a systematic map (Vendl et al. 2021 - accepted preprint [link](#)). The benchmark set we used consists of 10 papers.

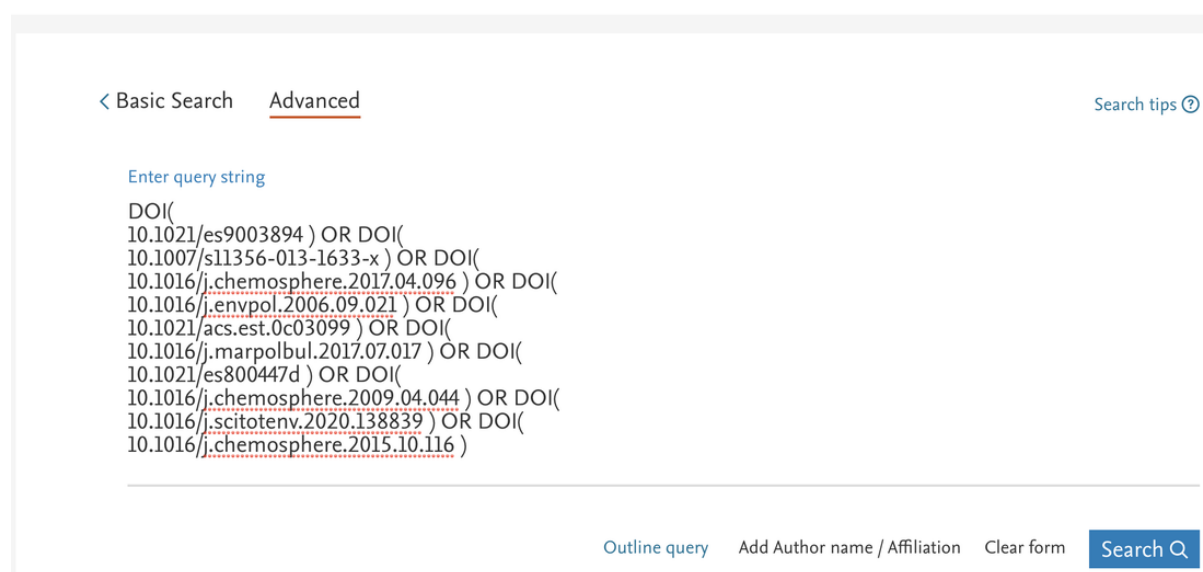
All steps are performed on Scopus using "Advanced document search", which allows us to use complex custom search strings but also has a useful feature of tracking history of recent searches (as a table at the bottom of the search page).

Here is the process in seven simple steps, with screenshots:

### 1.

Find your benchmark papers on Scopus using a search string made of paper's doi (search strings based on titles are longer and less accurate):

## Advanced search



The screenshot shows the Scopus search interface. At the top, there are tabs for "Basic Search" and "Advanced", with "Advanced" selected. A "Search tips" link is visible in the top right. Below the tabs, there is a section titled "Enter query string" containing a long search string of DOIs: `DOI( 10.1021/es9003894 ) OR DOI( 10.1007/s11356-013-1633-x ) OR DOI( 10.1016/j.chemosphere.2017.04.096 ) OR DOI( 10.1016/j.envpol.2006.09.021 ) OR DOI( 10.1021/acs.est.0c03099 ) OR DOI( 10.1016/j.marpolbul.2017.07.017 ) OR DOI( 10.1021/es800447d ) OR DOI( 10.1016/j.chemosphere.2009.04.044 ) OR DOI( 10.1016/j.scitotenv.2020.138839 ) OR DOI( 10.1016/j.chemosphere.2015.10.116 )`. At the bottom of the search box, there are buttons for "Outline query", "Add Author name / Affiliation", "Clear form", and a "Search Q" button.

### 2.

This should find all benchmark papers that are indexed in the database (unless associated doi is wrong, which sometimes happens). If some papers do not get picked up, search for them individually using title, then author or citing papers, get their "correct" doi or other identification code. In the example, all benchmark papers were found:

Search Sources Lists SciVal ↗

## 10 document results

DOI ( 10.1021/es9003894 ) OR DOI ( 10.1007/s11356-013-1633-x ) OR DOI ( 10.1016/j.chemosphere.2017.04.096 ) OR DOI ( 10.1016/j.envpol.2006.09.021 ) OR DOI ( 10.1021/acs.est.0c03099 ) OR DOI ( 10.1016/j.marpolbul.2017.07.017 ) OR DOI ( 10.1021/es800447d ) OR DOI ( 10.1016/j.chemosphere.2009.04.044 ) OR DOI ( 10.1016/j.scitotenv.2020.138839 ) OR DOI ( 10.1016/j.chemosphere.2015.10.116 )

[Edit](#) [Save](#) [Set alert](#)

### 3.

Now run the search you would like to be validated against the benchmark set, e.g.:

Enter query string

```
(TITLE-ABS-KEY(( *fluoroalkyl* OR polyfluo* OR perfluo* OR AFFF OR PFAS OR PFBA OR PFPeA OR PFHxA OR PFHpA OR PFOA OR PFNA OR PFDA OR PFDeA OR PFDcA OR PFUnDA OR PFUnA OR PFUA OR PFUDa OR PFDoA OR PFDODA OR PFTrDA OR PFTriDA OR PFTra OR PFTA OR PFTeDA OR PFO4DA OR PFO5DoDA OR PFBS OR PFBuS OR PFPeS OR PFHxS OR PFHpS OR PFOS OR PFNS OR PFDS OR FTS OR FTSA OR PFECHS OR FOSA OR PFOA OR NETFOSAA OR "Et-PFOA-AcOH" OR NMeFOSAA OR "Me-PFOA-AcOH" OR ADONA OR "Cl-PFAES" OR "F-53B" OR GenX OR "HFPO-TA" OR "Hydro-Eve" OR "Nafion BP2" OR POPS OR TEFLON OR TEFAL OR C8) AND (*concentr* OR bioaccumul* OR biomagnificat*) AND (wild* OR animal* OR fish* OR *bird* OR avian OR aves OR mammal* OR amphib* OR reptil* OR cetacean* OR *vertebrate* OR predator* OR prey* OR insect* OR crustacea* OR athropod* OR annelid* OR *worm* OR coral* OR pisc* OR frog* OR mollusc* OR fly* OR larva* OR nematod* OR cnidaria* OR anthozoa*) AND NOT (laboratory OR rats OR zebrafish* OR clinical* OR patient* OR worker* OR mouse OR mice OR vitro OR dairy OR farm* OR "cow milk" OR pig* OR dog* OR men OR women OR child* OR trial* OR medic* OR membrane* OR diabet* OR microb*)) ) AND ( LIMIT-TO ( EXACTKEYWORD , "Animals" ) OR LIMIT-TO ( EXACTKEYWORD , "Nonhuman" ) OR LIMIT-TO ( EXACTKEYWORD , "Animal" ) )
```

Outline query Add Author name / Affiliation Clear form [Search Q](#)

**4.**

This search finds over 2,000 references:  
2,101 document results

```
(TITLE-ABS-KEY(( *fluoroalkyl* OR polyfluo* OR perfluo* OR afff OR pfas OR pfba OR pfpea OR pfhxa OR pfhpa OR pfoa OR pfna OR pfda OR pfdea OR pfdca OR pfunda OR pfuna OR pfua OR pfuda OR pfdoa OR pfdoda OR pfttrda OR pfttrida OR pfttra OR pfta OR pfteda OR pfo4da OR pfo5doda OR pfbs OR pfbus OR pfpes OR pfhs OR pfhps OR pfos OR pfns OR pfds OR fts OR ftsa OR pfechs OR fosa OR pfoa OR netfosaa OR "Et-PFOA-AcOH" OR nmefosaa OR "Me-PFOA-AcOH" OR adona OR "Cl-PFAES" OR "F-53B" OR genx OR "HFPO-TA" OR "Hydro-Eve" OR "Nafion BP2" OR pops OR teflon OR tefal OR c8) AND (*concentr* OR bioaccumul* OR biomagnificat*) AND (wild* OR animal* OR fish* OR *bird* OR avian OR aves OR mammal* OR amphib* OR reptil* OR cetacean* OR *vertebrate* OR predator* OR prey* OR insect* OR crustacea* OR athropod* OR annelid* OR *worm* OR coral* OR pisc* OR frog* OR *mammal* )
```

Edit Save Set alert

**5.**

When you go back to the main Advanced Search window (it is easiest by clicking on "Search" link in the top right of the results window), below the box for specifying the searches, there is a table displaying your search history. On the top right of the table, there is a small box, titled "Combine queries...", that allows to quickly combine searches (you could also do that manually in the search box, but this requires careful copying and pasting long search strings). It is easy to combine search numbers from search history (in this case searches #5 AND #6):

Search history	Combine queries... <small>e.g. #1 AND NOT #3</small>	
	#5 AND #6	
6 (TITLE-ABS-KEY(( *fluoroalkyl* OR polyfluo* OR perfluo* OR afff OR pfas OR pfba OR pfpea OR pfhxa OR pfhpa OR pfoa OR pfna OR pfda OR pfdea OR pfdca OR pfunda OR pfuna OR pfua OR pfuda OR pfdoa OR pfdoda OR pfttrda OR pfttrida OR pfttra OR pfta OR pfteda OR pfo4da OR pfo5doda OR pfbs OR pfbus OR pfpes OR pfhs OR pfhps OR pfos OR pfns OR pfds OR fts OR ftsa OR pfechs OR fosa OR pfoa OR netfosaa OR "Et-PFOA-AcOH" OR nmefosaa OR "Me-PFOA-AcOH" OR adona OR "Cl-PFAES" OR "F-53... View More	2,101 document results	
5 DOI (10.1021/es9003894) OR DOI (10.1007/s11356-013-1633-x) OR DOI (10.1016/j.chemosphere.2017.04.096) OR DOI (10.1016/j.envpol.2006.09.021) OR DOI (10.1021/acs.est.0c03099) OR DOI (10.1016/j.marpolbul.2017.07.017) OR DOI (10.1021/es800447d) OR DOI (10.1016/j.chemosphere.2009.04.044) OR DOI (10.1016/j.scitotenv.2020.138839) OR DOI (10.1016/j.chemosphere.2015.10.116)	10 document results	

Top of page

**6.**

After pressing Enter the search is re-run with combined search strings:

## 10 document results

DOI (10.1021/es9003894) OR DOI (10.1007/s11356-013-1633-x) OR DOI (10.1016/j.chemosphere.2017.04.096) OR DOI (10.1016/j.envpol.2006.09.021) OR DOI (10.1021/acs.est.0c03099) OR DOI (10.1016/j.marpolbul.2017.07.017) OR DOI (10.1021/es800447d) OR DOI (10.1016/j.chemosphere.2009.04.044) OR DOI (10.1016/j.scitotenv.2020.138839) OR DOI (10.1016/j.chemosphere.2015.10.116)

[Edit](#) [Save](#) [Set alert](#)

### 7.

When the number of returned record equals the number of papers in the benchmarking set (as in this example) - all benchmark papers were found by the search string being validated. If the number is less - some benchmark papers were not found. Although you do not have to aim for 100% search string sensitivity, it is desirable to be able to locate most of the benchmark papers, so carefully revise your string and benchmark it again.

#### **Advantages of benchmarking using online database:**

- quick and accurate, because databases use their internal records IDs to check overlaps of the lists of references, rather than trying to match titles
- easily works with searches that bring up very large numbers of references - no need to download large files with all the records (there are usually limits on downloads - 2,000 records for Scopus, 1,000 for WoS, per download file)
- displays your search history, you can then export/copy as documentation of your benchmarking process (add your comments to this).