Abstract Sifter User Guide, Version 7.0

Availability: The Abstract Sifter and documentation is freely available for download here:

https://comptox.epa.gov/dashboard/downloads

Contact: Nancy Baker at baker.nancy@epa.gov

Disclaimer: The views expressed in this user guide are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Abstract Sifter User Guide – Version 7.0

This user guide describes the functionality of the PubMed Abstract Sifter. The reader is invited to download the tool from here: https://comptox.epa.gov/dashboard/downloads

This document provides guidance on the use of the Abstract Sifter through a series of screen shots showing the most common tasks in the tool followed by some helpful tips.

What's new in version 7?

Version 7 has one new major function and a few smaller nice-to-have things.

The new major feature is called MeSH mining and what it does is extract the MeSH terms for substances (chemicals, genes, proteins, etc.) from the MeSH annotations that are part of most PubMed records. There's a whole section walking through the functionality below.

Sheets overview

Let's start!

First open the Abstract Sifter file AbstractSifter_v7.xlsm. A security warning may appear. If so, be sure to enable content as shown in Figure 1.

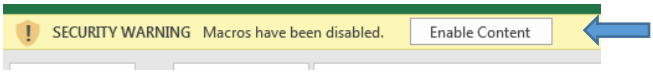


Figure 1. Enable macros upon opening

Once open you will see that the Abstract Sifter Excel file consists of 12 sheets. Each sheet is described briefly in the table below and covered in more detail in the body of this guide.

Sheet name	**Sheet Function**
ReadMe	Basic information on the sifter with links to more documentation
Main	Starting point for PubMed queries and for sifting
Abstract	The sheet where the citation abstract is shown
Notes	Notes and tags are inserted here
Log	Log of every query run on Main sheet
PathwayQueries	Repository of sample queries to use in research disease or treatment pathways (e.g., AOPs)
SampleQueries	Repository of sample queries to use as starting points
Landscape	High level view of literature for entities

CuratedLists	A place to keep lists of chemicals or other entities
TermExpand	Mapping term feature sets
MeSHMine	New in Version 7! This is where MeSH substance terms are written.
Archive	*This only exists if you use the Archive function, available through a button on the Main sheet*
Summary	*This only exists if you use the Summarize function, available through a button on the Main sheet*

A. Main Sheet

The Main sheet is where the basic functionality of Abstract Sifter occurs, including functions we call "sifting". To begin using the Abstract Sifter, the end-user clicks on the Run a query button at the top of the screen in the Main sheet. A form is displayed on which the user enters a query. In the example, we are showing a very simple query: "fipronil", but these queries can be more complex. (Figure A1) The text that the user enters into the box is sent to PubMed, so all PubMed syntax rules apply. (For a review of this syntax, visit https://www.ncbi.nlm.nih.gov/books/NBK3827/) NOTE: NLM changed PubMed early 2020 and reprogrammed the search algorithms. Since then, the search algorithms used by the web services do not return the exact counts returned by the PubMed web site.

Note also, that the end-user can optionally append the result of a query to the records currently on the Main sheet.

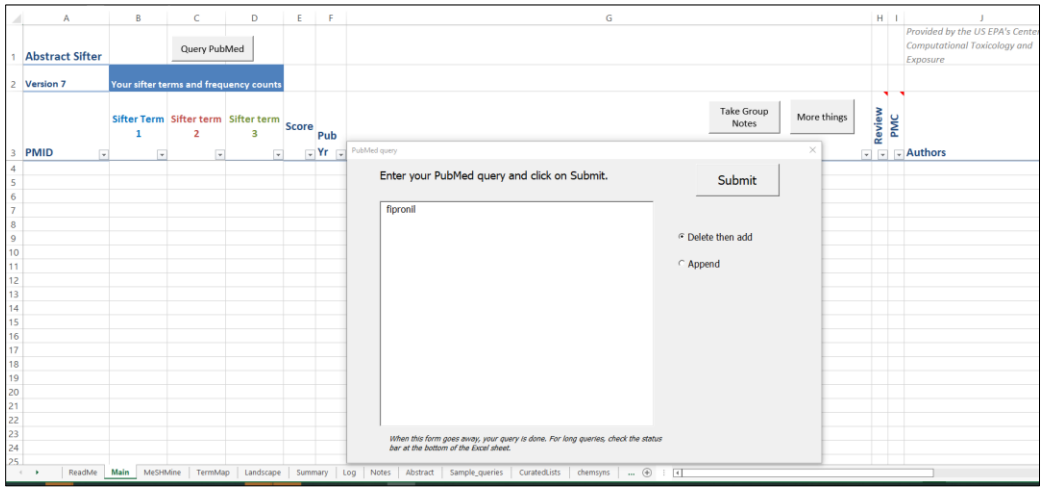


Figure A1. Running a PubMed query

When finished entering the query, the user clicks on Submit and the query is packaged by a Visual Basic Application (VBA) into an e-utility command that is passed to the NCBI (National Center for Biotechnology Information) web services. (Note that using Sifter Query PubMed capability requires internet access.) The first response returned by the utility is the number of articles found. (Figure A2) This number is displayed, and the user is asked if he/she want to continue. If the number of articles is

over 10,000, the query will not be run and the user is encouraged to refine the query to return fewer records.

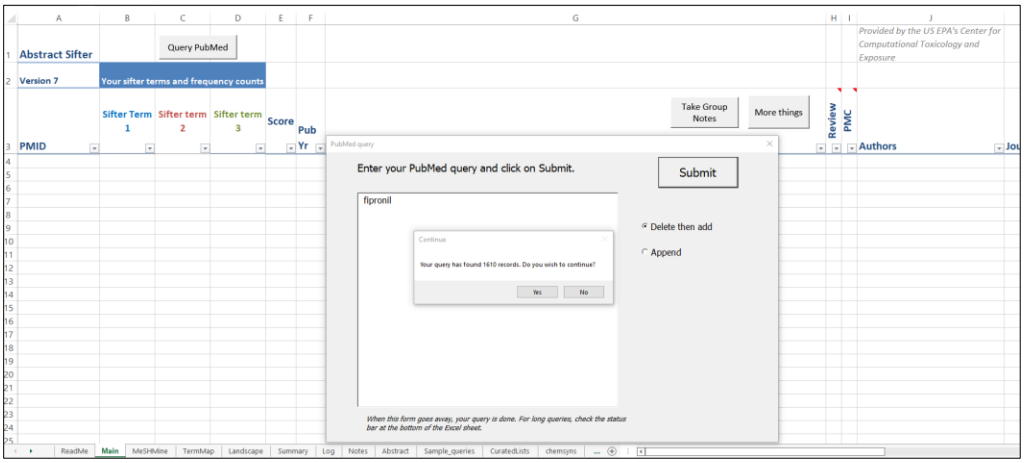


Figure A2. Responding to PubMed

If the returned results are fewer than 10,000 and the user indicates he/she wants to continue, the articles are downloaded from NCBI by Excel, and regular expressions are used to parse the citations for title, abstract, authors, publication year, journal, and PubMed identifier. Each record returned is inserted into a row in the Main sheet. Any rows in the Main sheet from a previous query are deleted unless the end-user chose the Append option. The Append option adds the new results to the end of the Main sheet. You can watch the status bar at the bottom of Excel to see how far along the retrieval process is. For longer queries, a cancel form will appear and the user can use it to cancel the process.

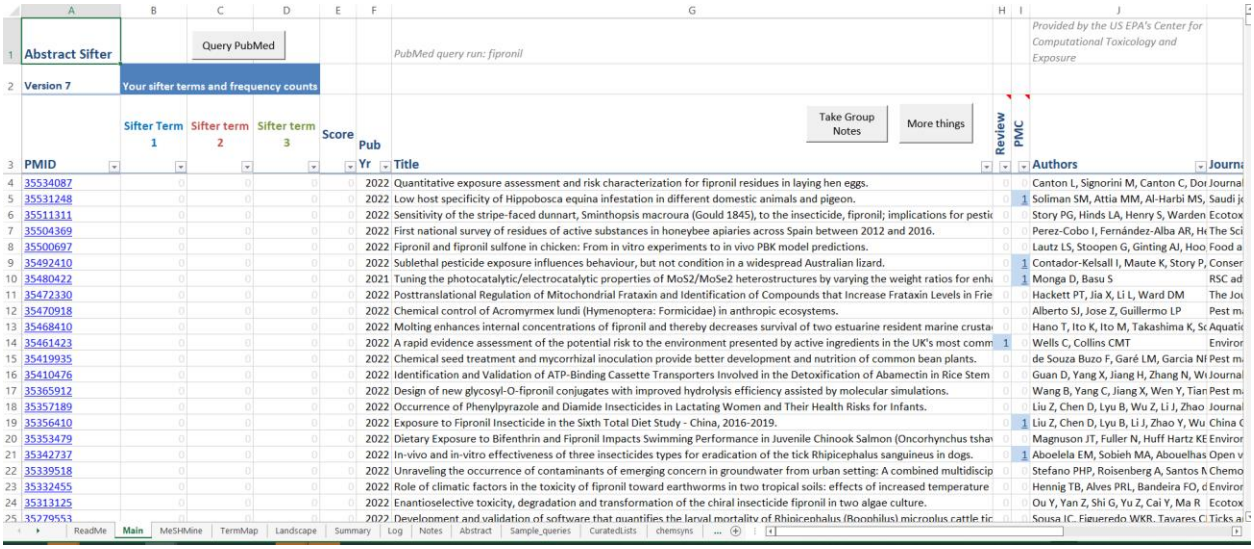


Figure A3. Results from PubMed query - before sifting

At this point the results of the query are stored in the Main sheet and can be browsed like any other data in a spreadsheet (Figure A3); however, the most effective way to find articles of interest is to use

the innovative sifter functionality. To demonstrate this functionality, we will continue to use our example of fipronil.

Let us suppose at this point that we are looking for dose-response toxicity data for fipronil. We type the term “fipronil” in cell B3, “toxic” in C3, and “mg” in D3. As we finish typing and move to the next cell, the Abstract Sifter will count the occurrences of the terms in the title, abstract, and key words combined. The citations can then be sorted by these counts, either individually or by the total. Figure A4 shows what the Sifter looks like when these terms have been entered into cells B3, C3, and D3 and then the entries sorted by occurrence counts of “fipronil” in descending order. PubMed article 32723848 has 28 occurrences of “fipronil”, 4 of “toxic”, and none of “mg/kg”. Looking farther down, PubMed article 12442503 has 21 of fipronil, 13 of “toxic” and four of “mg”. This article indeed describes a toxic doses of the chemical in various animal species. The sifter terms can be changed as many times as you want. Try “ppm” instead of “mg” or try “ticks”, “fleas”, “cats”, “dogs”.

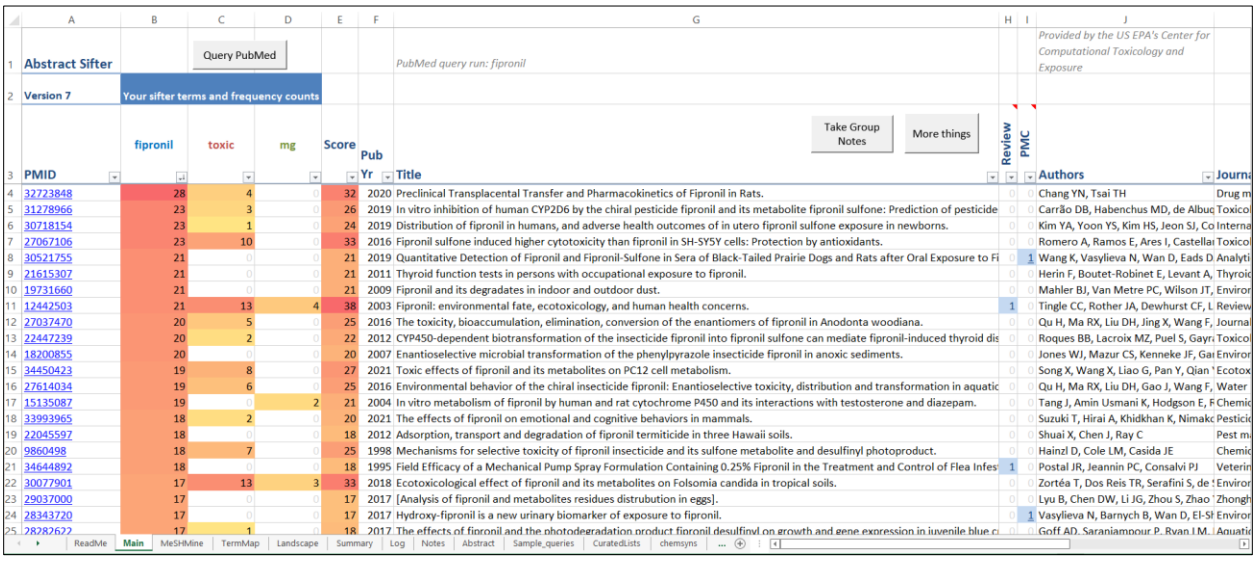


Figure A4. After sifter terms were entered into cells B3, C3, D3 and sorting on B3

Sifting the results through specifying sifter terms in B3, C3, and D3 can be repeated as many times as the user wishes. Similarly, new PubMed queries can be run, altered, rerun.

Let’s look at the other buttons on the Main sheet. The Take Group Notes button is discussed in section C of this guide. The More Things button contains some other actions that are often useful. Let’s check that out. Click on the More Things button reveals the menu displayed in Figure A5.

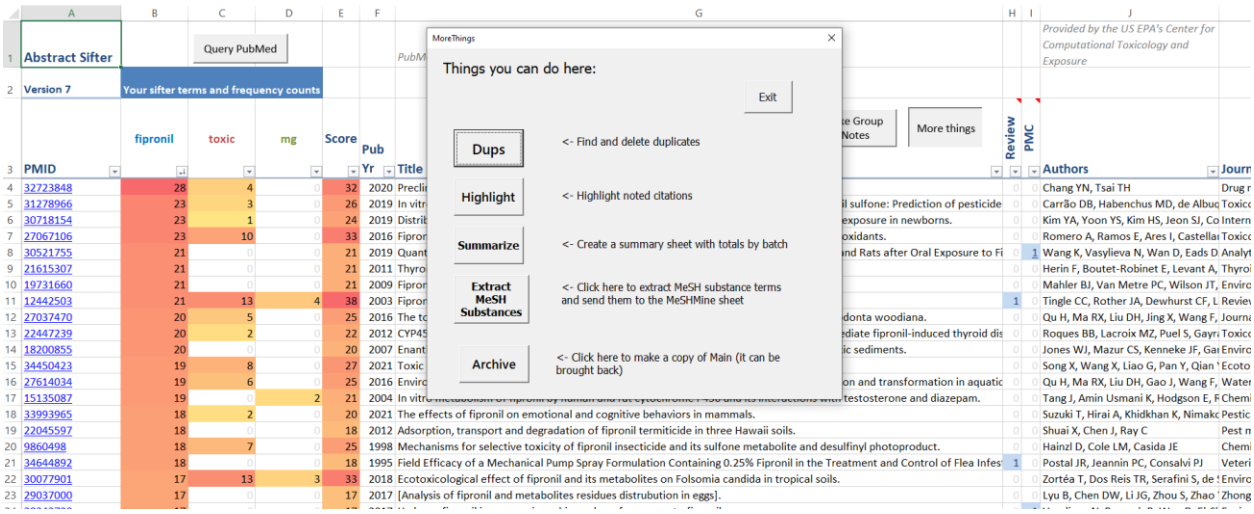


Figure A5. The More Things button on the Main sheet

The Dups button will find and select duplicate records and ask if you want to delete them. Let's say you run a query "fipronil AND fleas" and then run a query "fipronil AND ticks" and append the results for the second query to the first. There are bound to be duplicates. In this example, 39 duplicates were found and deleted shown in Figure A6.

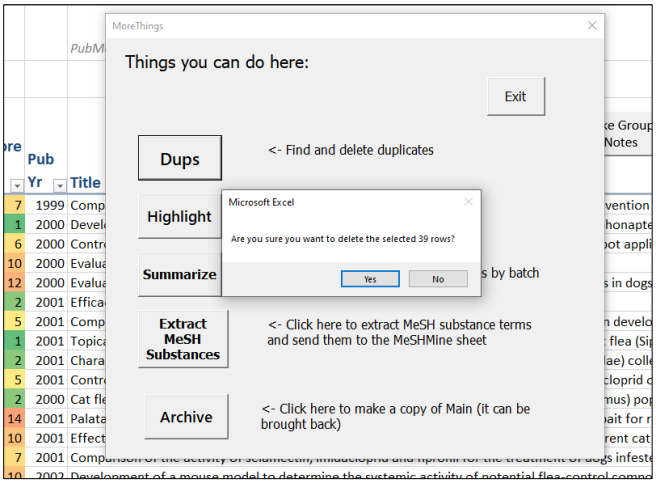


Figure A6. Deleting duplicates

The Highlight button will be covered in the Notes section of this manuscript. The Summarize function will be covered in the Log / Batch section. Let look at the Archive button next.

Sometimes people work for hours to retrieve a set of PubMed citations through a complex set of queries. The Main sheet is then the result of a lot of work. It used to be that running another query would clear the Main sheet and some users were hesitant to do this. The Archive button will copy the records on the Main sheet to another sheet called Archive. Once that is done, the user can run any number of queries sending results to the Main sheet and then when done can go to the Archive sheet and send back the contents to Main.

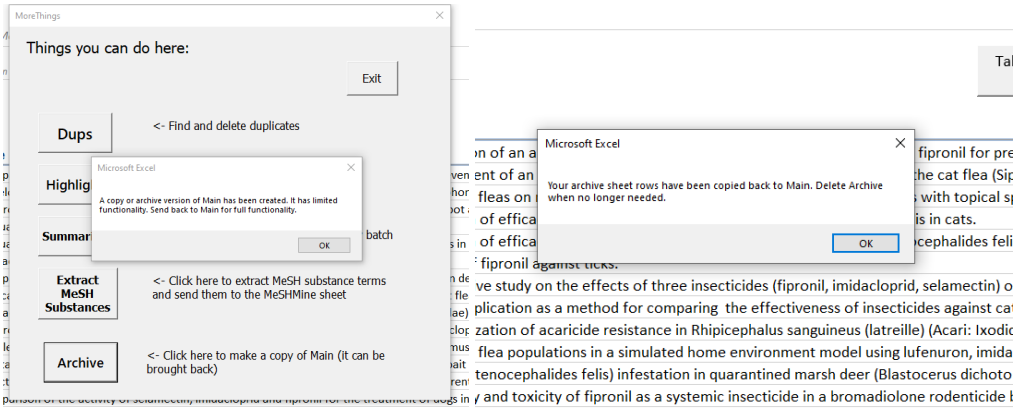


Figure A7. Archiving and unarchiving the results on the Main sheet.

The other option on the More Things menu is Extract MeSH Substances. This is an powerful new feature in Version 7. See section H below for a step-by-step description.

B. Abstract Sheet

To see the abstract for any of the retrieved articles, we can either click on the PubMed ID hyperlink to be taken to PubMed, or we can double-click on any other cell in the row for this article. This action brings us to the Abstract sheet where the abstract is displayed along with other article meta-data like title and authors (Figure B1).

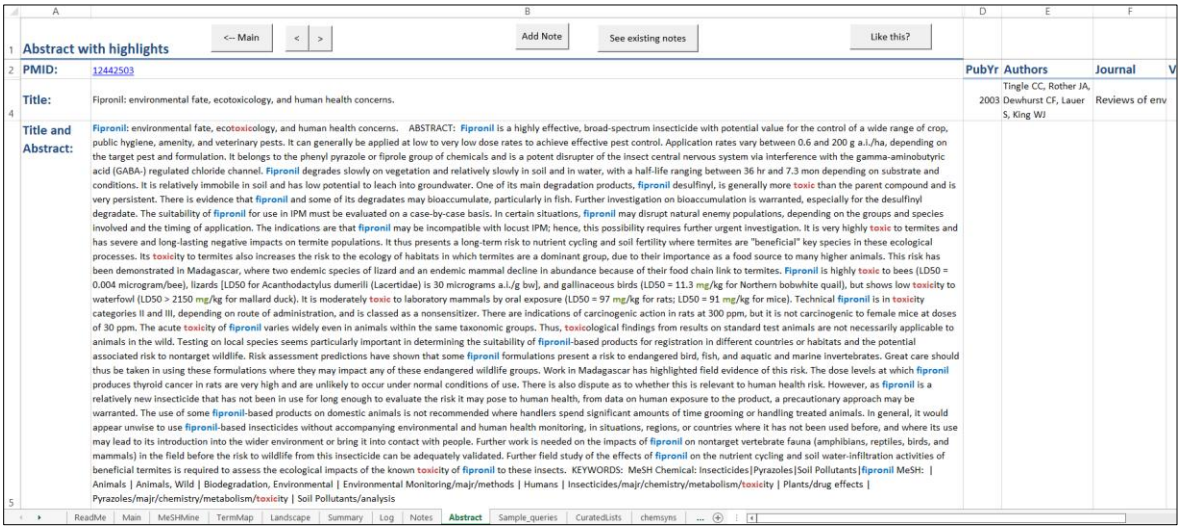


Figure B1. An abstract with highlighted sifter terms. Fipronil is blue, toxic is red, and mg is green.

There are several aspects of the Abstract sheet that are important to note. First, the sifter terms in the abstract are highlighted. The font colors reflect the colors of the fonts in cells B3, C3, and D3 entered into the Main Sheet. This highlighting makes the reading the abstract easier by drawing attention to sentences that might be of more interest.

Let's go through the other buttons on the Abstract sheet. The Like this? Button allows the user to find articles in PubMed that are similar to the article on the Abstract sheet or find articles in PubMed Central

that cite that article in question. (Figure B2.) These functionalities are likely familiar as they are offered on the PubMed Entrez web site. The results can be appended to results already on the Main sheet. A colorization feature has been added to make the titles either yellow or green. This way the user can see which rows were from the original query and which were retrieved from the Like this functionality.

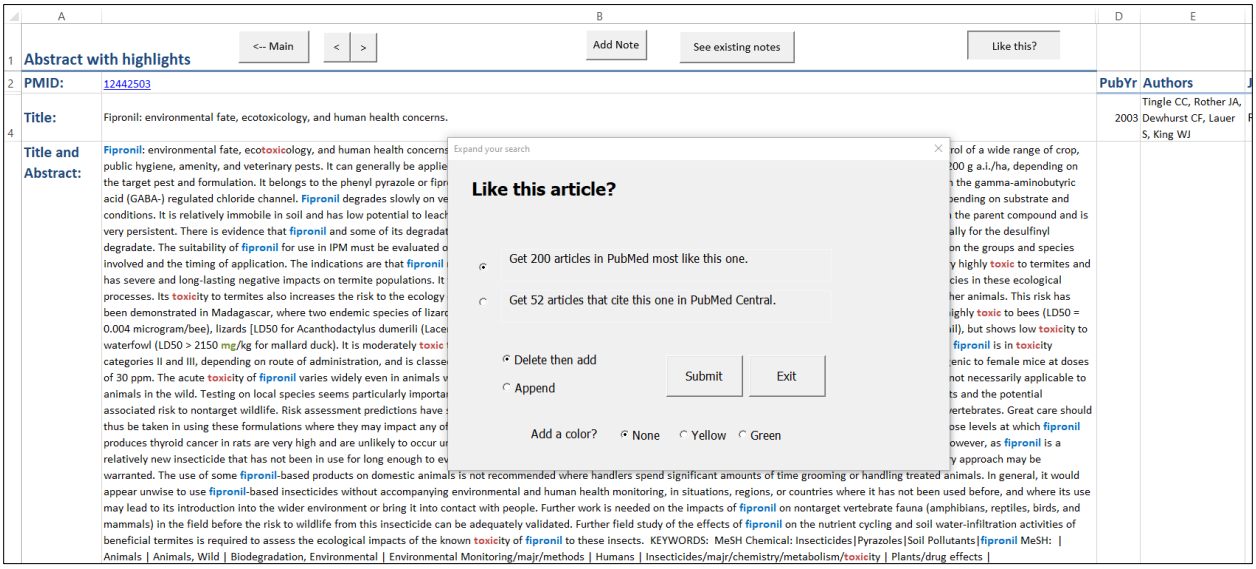


Figure B2. Clicking on the Like this? Button on the Abstract Sheet.

If you have already taken a Note on this article you can click on the See Existing Notes button to be taken to that citation on the Notes sheet. If no note is yet taken, a message saying that will be displayed. The Add Note button brings up the Notes form and you can enter a Note.

C. Notes Sheet

Given the dynamic nature of the Sifter, many users find it helpful to be able to make notes on articles that they want to track. There are two ways using the Sifter to take notes: one way is through the Main sheet, and the other way starts with the Abstract sheet. To return to our case study, let us say that we have found a set of articles on the Main sheet that we know we need to read in depth. We can select these articles and then click on the Take Group Notes button. A form appears where we can enter information into fields called Tag and Notes. These elements are self-defined. We can also click on yes, no, or maybe. The note-taker can enter her/his initials or name in the Who field. This set of variables is a quick way to associate articles with a note. Notice that these choices each come with a color (yes-green, no-red, and maybe-yellow). Entering any of these fields is optional. (Figures C1 and C2) When we click on the OK button, each article selected will be inserted into the Notes page with the corresponding

information (Figure C3) and the PubMed ID (PMID) on the Main sheet will be colorized.

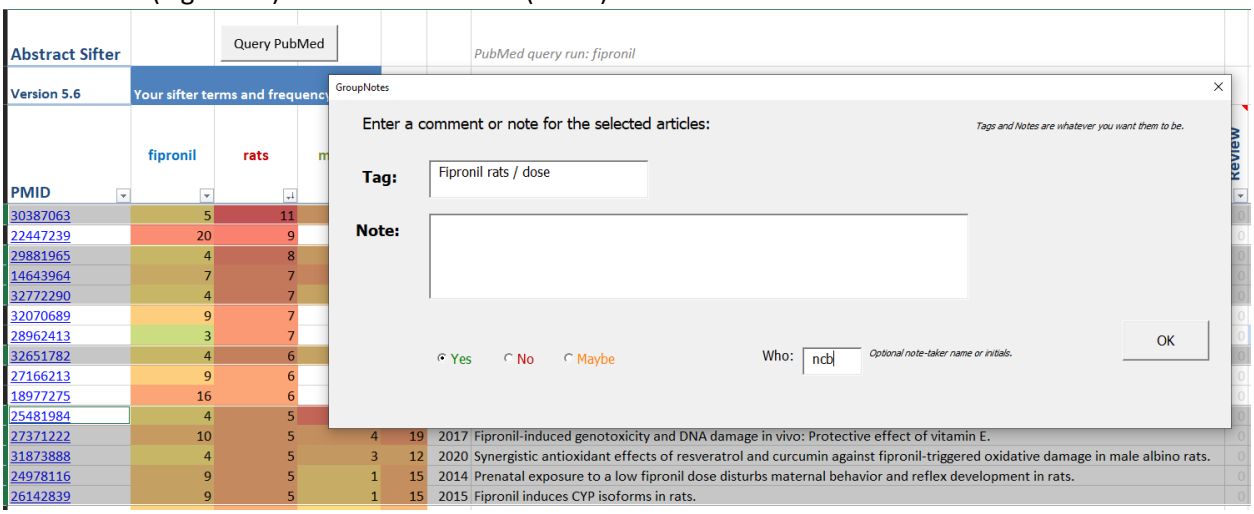


Figure C1. Taking group notes

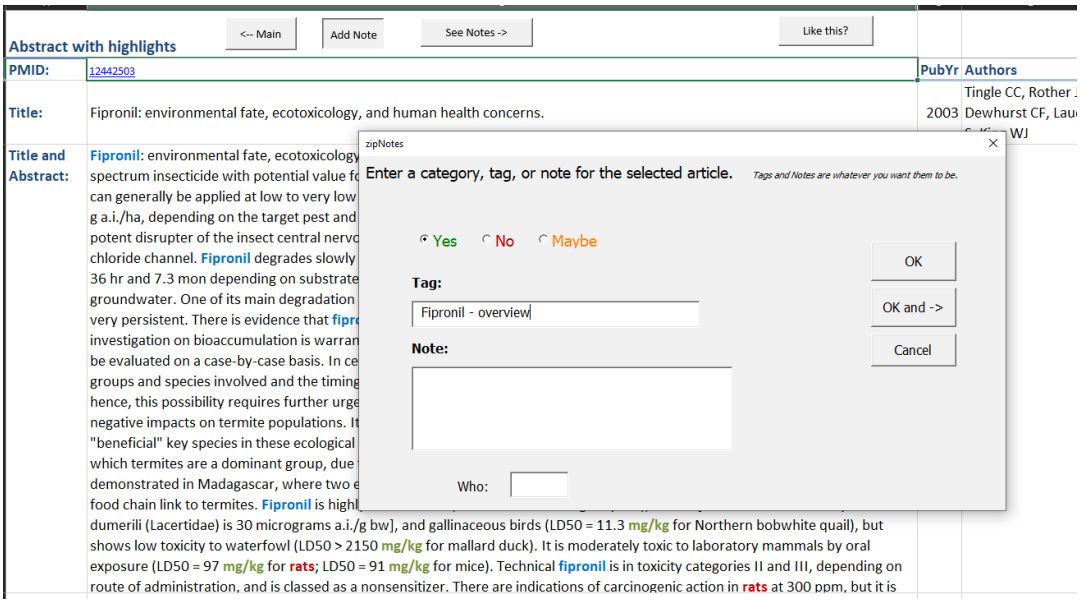
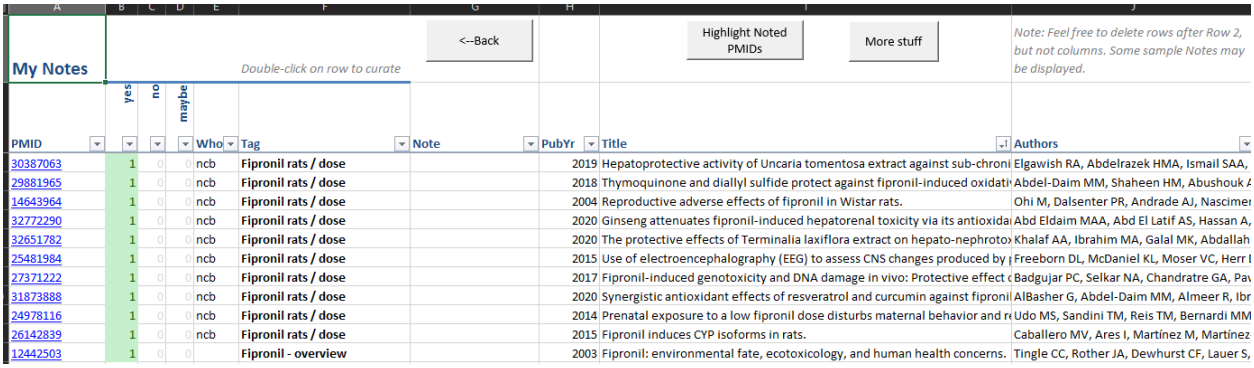


Figure C2. Taking single notes on the Abstract sheet



C3. The Notes sheet. Be sure to save your workbook.

The second option for note taking starts with the Abstract Sheet. (Figure C2) The "Add Note" button in the top row allows notes to be inserted into the Notes Sheet.

The note-taking can be used to help keep track of which citations have been read and evaluated and which have not. On the Main sheet the PMIDs can be sorted by the noted color using the built-in Excel sorting functionality. (Figure C4) The More Things button on the Main sheet will allow the end-user to delete duplicates and highlight noted citations with colorization in the cases where you've run some queries and the colors need to be refreshed.

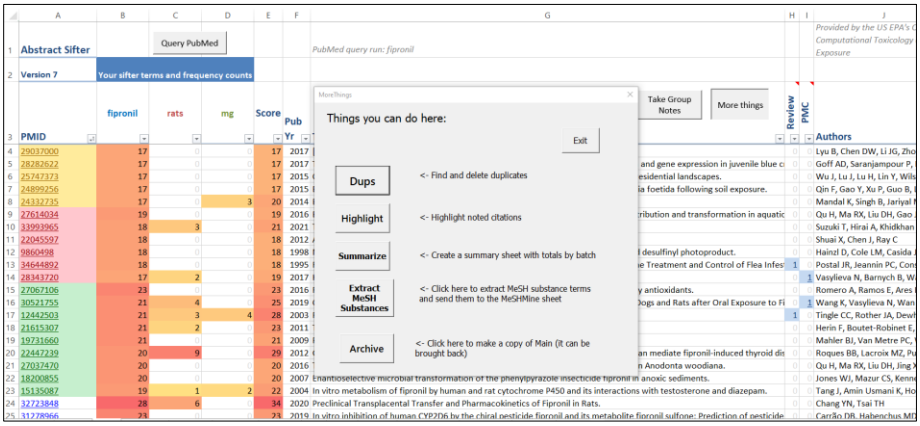


Figure C4. After clicking on More things, then and then sorting by color

The user can make changes to the Notes sheet by editing, adding or deleting rows below row 2. Double-clicking on a row in the Notes sheet brings up a curation form (Figure C5). Here the end-user can enter and update tags, notes. Text in the large box can be selected and dragged to any of the smaller tag and notes boxes. The Refresh button brings the title and abstract back in and refreshes the view. The purpose of this form is to allow quick easy extraction of information from the abstract into separate fields. The copy button sends the elements of the citation and your notes to the clipboard; then you can open up Word and paste the clipboard contents there. The Notes sheet can be copied, printed, or the Notes can be exported to a text file for further import into other tools like Word.

New in this version is the capability to attach a pdf. Before you can attach a pdf, the file must reside on your computer. Click the Attach Local PDF button and find the pdf. You can open the pdf by clicking on the Open button or on the pdf hyperlink in column W. Note that while Abstract Sifter files can be shared with colleagues, the pdfs are local and specific to one machine.

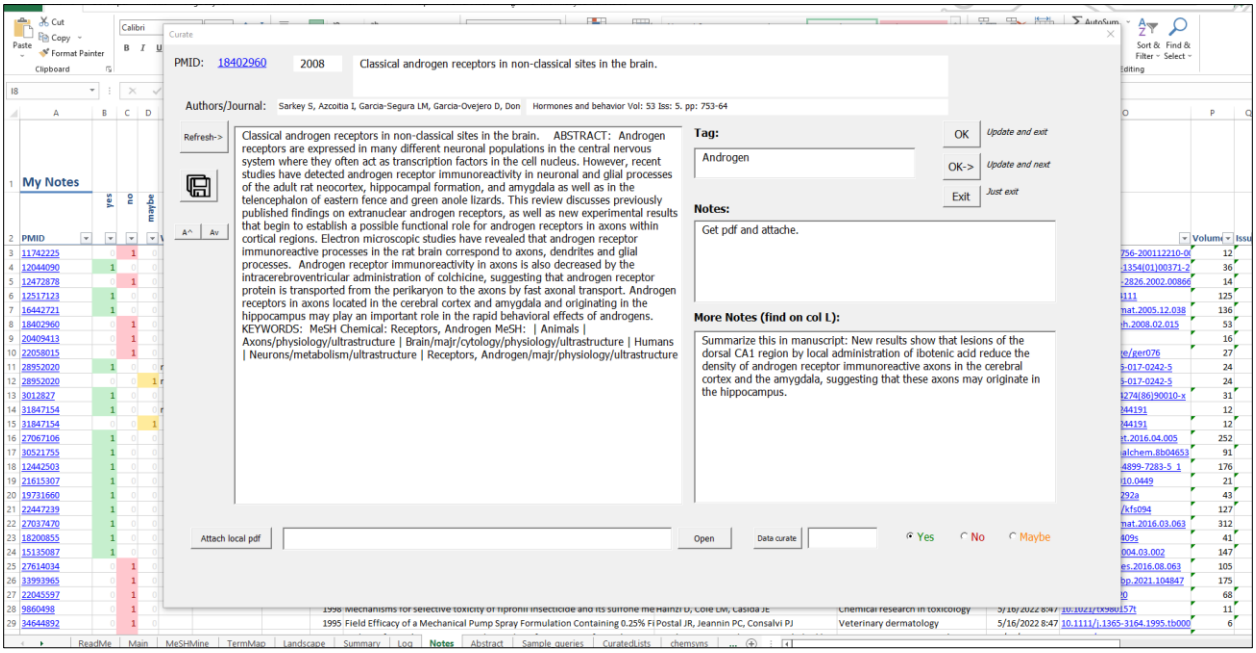


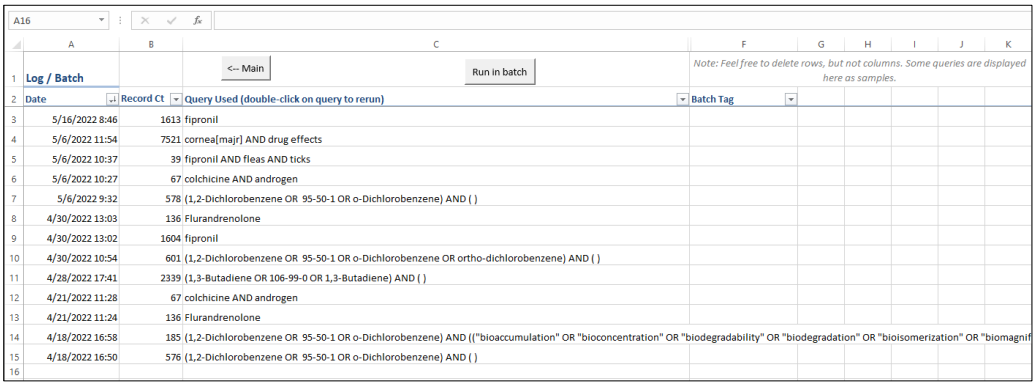
Figure C5. Double-clicking on a row in the Notes sheet brings up a curation form.

Later in this document there are tip sections that relate to note-taking. See Exporting to other applications from the Notes sheet.

D. Log / Batch sheet

The Log sheet keeps track of the queries you have run. The Abstract Sifter routines insert a row into the sheet every time you complete a query. These queries can be viewed and rerun. To rerun a query, simply double-click on it. (Figure D1) The Batch Run option allows the user to run multiple queries and append the results from each on the Main sheet. To run in batch, select rows and click on the Run in Batch button. A Batch Tag can be added or modified on the Log sheet. This tag will be added to the Main sheet results and used to help summarize the results of multiple runs. This summary functionality is accessible on the Main sheet through the More things button.

Delete any or all rows after Row 2 if you want to clear old entries.



	A	B	C	F
1	Log / Batch		<-- Main Run in batch	Note: Feel free to delete rows, but not columns. Some queries are displayed here as samples.
2	Date	Record Ct	Query Used (double-click on query to rerun)	Batch Tag
3	5/16/2022 8:46	1613	fipronil	
4	5/6/2022 11:54	7521	cornea[majr] AND drug effects	
5	5/6/2022 10:37	39	fipronil AND fleas AND ticks	
6	5/6/2022 10:27	67	colchicine AND androgen	
7	5/6/2022 9:32	578	(1,2-Dichlorobenzene OR 95-50-1 OR o-Dichlorobenzene) AND ( )	
8	4/30/2022 13:03	136	Flurandrenolone	
9	4/30/2022 13:02	1604	fipronil	
10	4/30/2022 10:54	601	(1,2-Dichlorobenzene OR 95-50-1 OR o-Dichlorobenzene OR ortho-dichlorobenzene) AND ( )	
11	4/28/2022 17:41	2339	(1,3-Butadiene OR 106-99-0 OR 1,3-Butadiene) AND ( )	
12	4/21/2022 11:28	67	colchicine AND androgen	
13	4/21/2022 11:24	136	Flurandrenolone	
14	4/18/2022 16:58	185	(1,2-Dichlorobenzene OR 95-50-1 OR o-Dichlorobenzene) AND (("bioaccumulation" OR "bioconcentration" OR "biodegradability" OR "biodegradation" OR "bioisomerization" OR "biomagnif	
15	4/18/2022 16:50	576	(1,2-Dichlorobenzene OR 95-50-1 OR o-Dichlorobenzene) AND ( )	
16				

Figure D1. View of the Log sheet

E. Landscape sheet

The Landscape sheet provides an overview of the literature to the user for a set of entities, for example, a list of chemicals or genes. Figure 12 shows an example of a Landscape sheet built by a researcher interested in the toxicity of a particular set of chemicals. Let's take a look at that first. Queries designed to find the chemicals of interest are entered into Column C and in this case, a short version of the chemical name is in Column B. The queries in Row 3 are typical ones used in searching for articles about different kinds of chemical toxicity. We will refer to these queries as subject matter queries. (Note: Column A on the Landscape sheet is often hidden. Go ahead and unhide it and use it when you have a DSSTox chemical identifier.)

The premise behind the design of the Landscape sheet is very simple: PubMed queries will be built by taking the values in Column C (in this example chemical names and corresponding CAS numbers) and appending this query text to the subject matter query text in Row 3 with an " AND " in between the two query parts.

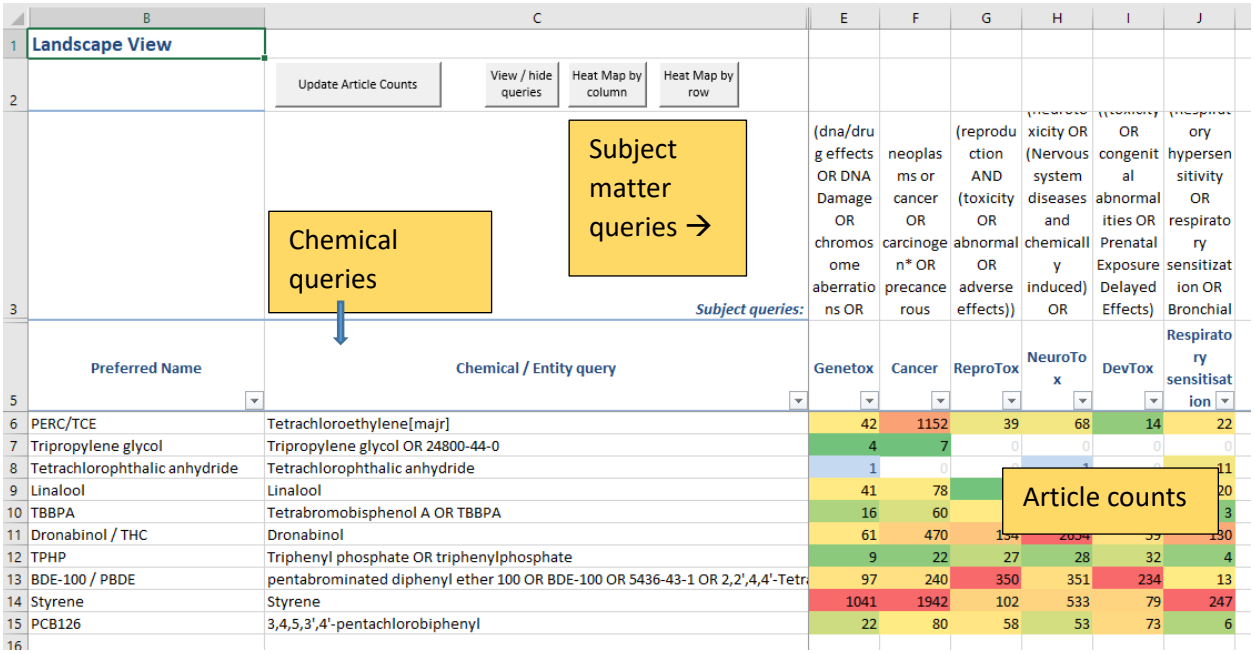
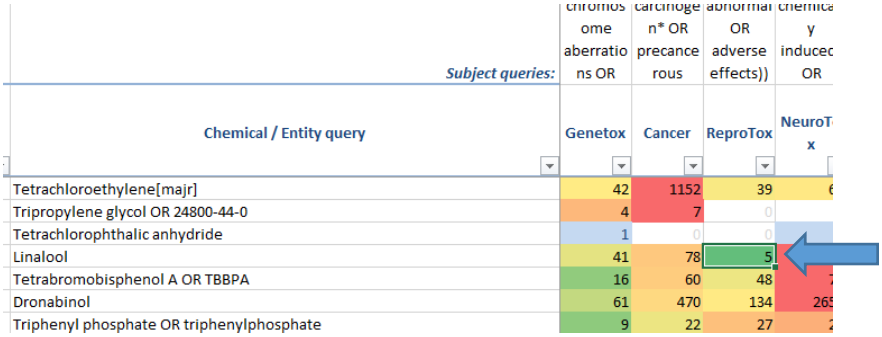


Figure E1. Structure of the Landscape sheet

To illustrate, we will double-click on the cell with the arrow pointer in Figure E2. When we double-click on this cell this tells the Abstract Sifter to take the query text in Column C about Linalool and append it to query text designed to find citations about reproductive toxicity. Figure E3 shows the constructed query. We can then click on Submit and the query gets sent to PubMed and we can then see the results on the Main sheet. The number of articles retrieved from PubMed is 5. That count is placed in the corresponding Landscape cell that we just clicked on.



Subject queries:	chromosome aberrations OR	carcinogen* OR precancerous	abnormal OR adverse effects))	chemicaly induced OR
Chemical / Entity query	Genetox	Cancer	ReproTox	NeuroTox
Tetrachloroethylene[majr]	42	1152	39	6
Tripropylene glycol OR 24800-44-0	4	7	0	
Tetrachlorophthalic anhydride	1	0	0	
Linalool	41	78	5	
Tetrabromobisphenol A OR TBBPA	16	60	48	7
Dronabinol	61	470	134	265
Triphenyl phosphate OR triphenylphosphate	9	22	27	2

Figure E2. Double-click on article count cells

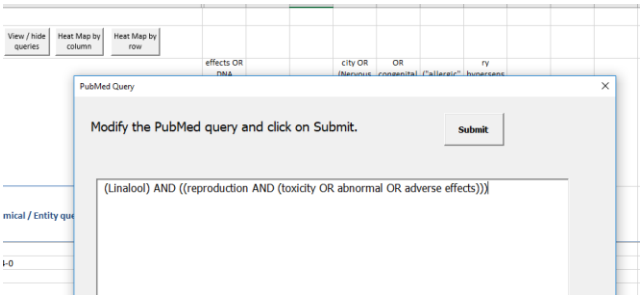


Figure E3. Constructed query

Now let's add to the Landscape sheet. Figure E4 shows how we added a new chemical to the list: aspirin. To find out the article counts for aspirin, select empty cells on the same row as aspirin, then click on Update Article Counts button. Excel will build each query from the aspirin part and the subject matter part and send each query to PubMed to find out how many citations satisfy the query. The article counts are placed in the corresponding cells. To run the query and retrieve the results, just double-click on any of the article count cells.

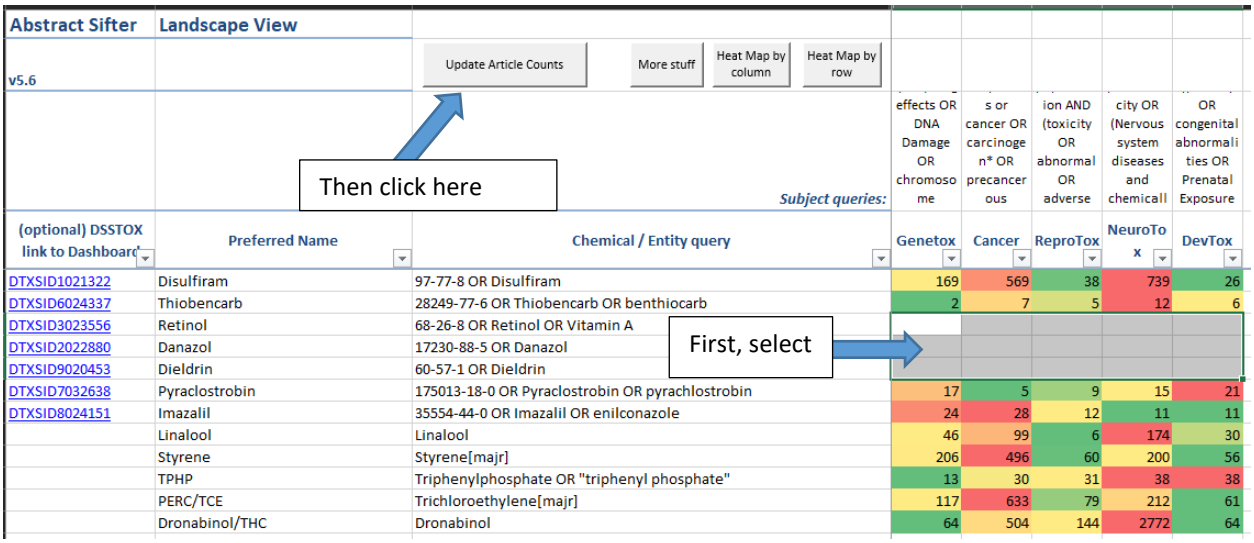


Figure E4. Adding results to the Landscape sheet

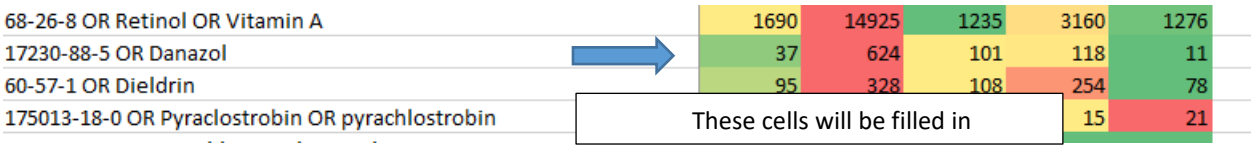


Figure E5. After clicking on Update Article Counts

New subject matter queries can be entered as well. The query part goes into Row 3 and a heading (of your choice) goes into Row 5. See the example below where the PubMed query part: skin OR dermatitis is entered with the heading skin. Next highlight the cells underneath and click on the Update Article Counts button. The counts of articles satisfying the queries are placed in the cells. What's happening behind the scenes? For each cell, a query is being built by the Abstract Sifter and sent to PubMed to retrieve a record count. That record count is then inserted into the corresponding cell. (Figure E6.)

Keep in mind that our examples revolve around chemicals, but that does not mean one is limited to chemicals. The entries in Column C and in Row 3 can be whatever you the end-user want them to be: genes, proteins, diseases, authors …

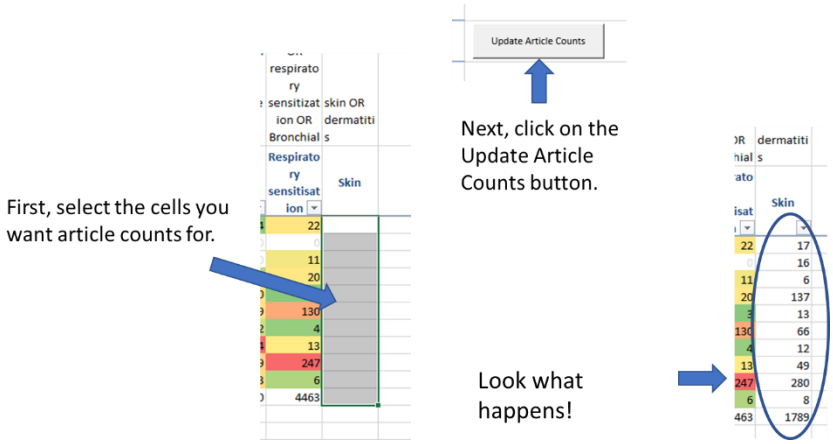


Figure E6. Steps for retrieving counts

Making things look good

The Landscape sheet has three buttons that make formatting easy (E7). The heat map buttons will quickly apply heat map coloring to the cells with article counts either by column or by row. Try them out!

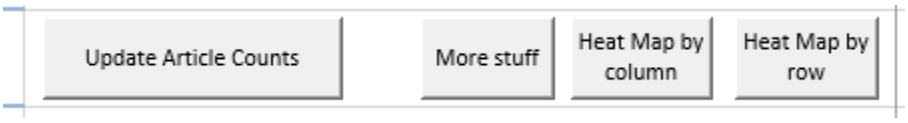
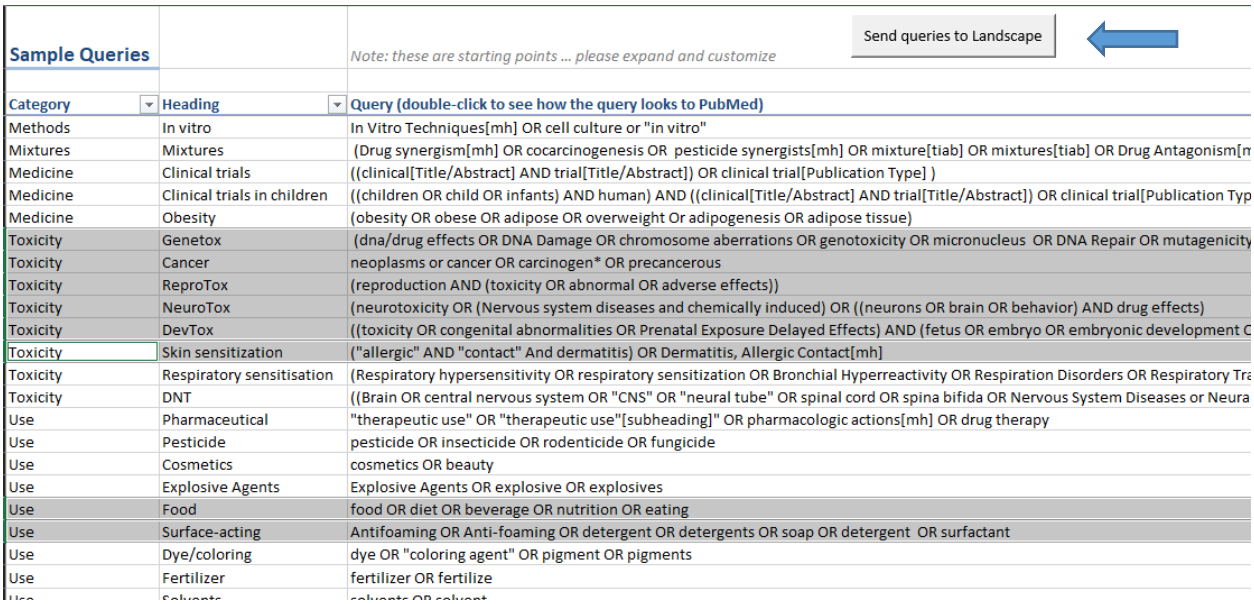


Figure E7. Buttons on the Landscape page include formatting actions.

F. Sample_Queries, Pathway_Queries, CuratedLists Sheets

The Sample_queries and pathway_queries sheets function in a very similar way. We will use the Sample_queries sheet as an example in this user guide. Both sheets contain a number of sample subject matter queries that the end user can use as a starting point for building a Landscape view of a set of entities. Let's see how. First, we will clean off the old subject matter queries by deleting columns E-L on the Landscape sheet. (You can let the previous work stay if you wish.) Next, on the Sample_Queries sheet we will select rows with queries of interest then we click on the button Send Queries to Landscape (Figure F1).



Sample Queries

Note: these are starting points … please expand and customize

Send queries to Landscape

Category	Heading	Query (double-click to see how the query looks to PubMed)
Methods	In vitro	In Vitro Techniques[mh] OR cell culture or "in vitro"
Mixtures	Mixtures	(Drug synergism[mh] OR cocarcinogenesis OR pesticide synergists[mh] OR mixture[tiab] OR mixtures[tiab] OR Drug Antagonism[m
Medicine	Clinical trials	((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trial[Publication Type])
Medicine	Clinical trials in children	((children OR child OR infants) AND human) AND ((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trial[Publication Typ
Medicine	Obesity	(obesity OR obese OR adipose OR overweight Or adipogenesis OR adipose tissue)
Toxicity	Genetox	(dna/drug effects OR DNA Damage OR chromosome aberrations OR genotoxicity OR micronucleus OR DNA Repair OR mutagenicity
Toxicity	Cancer	neoplasms or cancer OR carcinogen* OR precancerous
Toxicity	ReproTox	(reproduction AND (toxicity OR abnormal OR adverse effects))
Toxicity	NeuroTox	(neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR brain OR behavior) AND drug effects)
Toxicity	DevTox	((toxicity OR congenital abnormalities OR Prenatal Exposure Delayed Effects) AND (fetus OR embryo OR embryonic development C
Toxicity	Skin sensitization	("allergic" AND "contact" And dermatitis) OR Dermatitis, Allergic Contact[mh]
Toxicity	Respiratory sensitisation	(Respiratory hypersensitivity OR respiratory sensitization OR Bronchial Hyperreactivity OR Respiration Disorders OR Respiratory Tra
Toxicity	DNT	((Brain OR central nervous system OR "CNS" OR "neural tube" OR spinal cord OR spina bifida OR Nervous System Diseases or Neura
Use	Pharmaceutical	"therapeutic use" OR "therapeutic use"[subheading]" OR pharmacologic actions[mh] OR drug therapy
Use	Pesticide	pesticide OR insecticide OR rodenticide OR fungicide
Use	Cosmetics	cosmetics OR beauty
Use	Explosive Agents	Explosive Agents OR explosive OR explosives
Use	Food	food OR diet OR beverage OR nutrition OR eating
Use	Surface-acting	Antifoaming OR Anti-foaming OR detergent OR detergents OR soap OR detergent OR surfactant
Use	Dye/coloring	dye OR "coloring agent" OR pigment OR pigments
Use	Fertilizer	fertilizer OR fertilize

Figure F1. Selecting rows with queries of interest

Our Landscape sheet then looks like Figure F2.

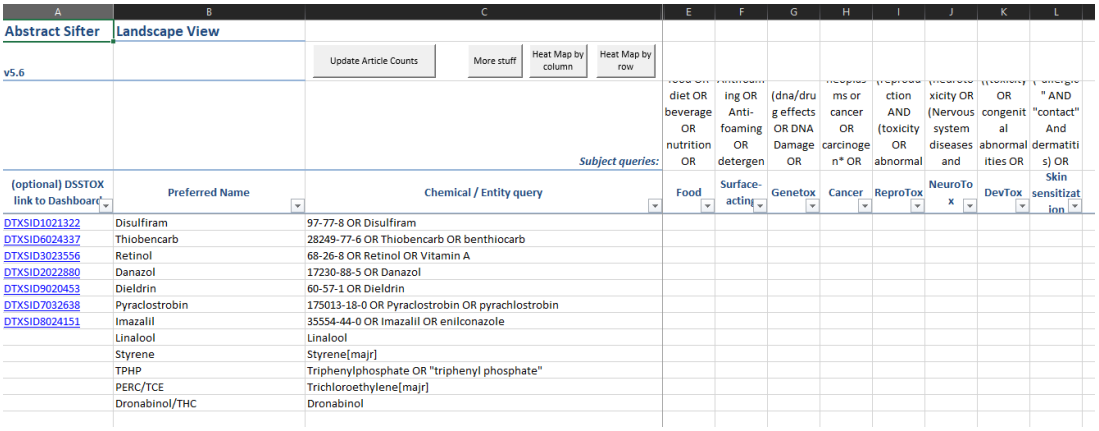


Figure F2. New queries on Landscape sheet

Next, we select the article count area and then click on *Update Article Counts*.

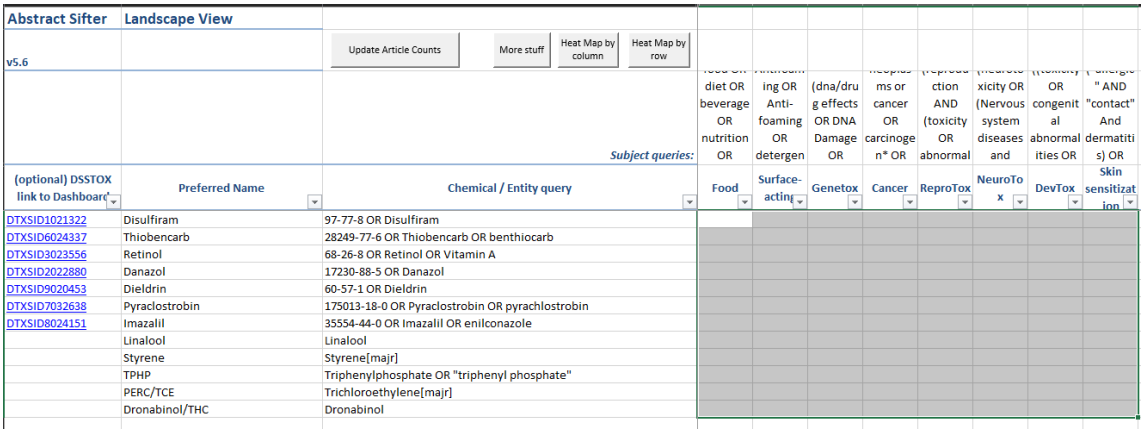


Figure F3. Selecting the cells for article counts

Once the article counts are populated, we click on Heat Map by Row and then on Hide queries. Our resulting Landscape view looks like Figure F4. To run the query and retrieve the results, just double-click on any of the article count cells.

Note that the Pathway_queries sheet works the same way: select the rows of interest and click on the button Send queries to Landscape. To add your own queries to either the pathway or sample queries sheets, just enter in new rows. Use the existing rows as a template to indicate where headers go and where the query text goes. Be mindful of parentheses and be sure to test the queries.

Abstract Sifter	Landscape View											
v5.6		Update Article Counts	More stuff	Heat Map by column	Heat Map by row							
					Subject queries:	diet OR beverage OR nutrition OR	ing OR Anti-foaming OR detergen	(dna/drug effects OR DNA Damage OR	ms or cancer OR carcinogen* OR	ction AND (toxicity OR abnormal	xicity OR (Nervous system diseases and	OR congenital abnormalities OR	" AND "contact" And dermatitis) OR
(optional) DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query				Food	Surface-acting	Genetox	Cancer	ReproTox	NeuroTox	DevTox	Skin sensitization
DTXSID1021322	Disulfiram	97-77-8 OR Disulfiram				308	29	169	569	38	739	26	25
DTXSID6024337	Thiobencarb	28249-77-6 OR Thiobencarb OR benthiocarb				9	0	2	7	5	12	6	1
DTXSID3023556	Retinol	68-26-8 OR Retinol OR Vitamin A				19770	585	1690	14925	1235	3160	1276	38
DTXSID2022880	Danazol	17230-88-5 OR Danazol				94	43	37	624	101	118	11	0
DTXSID9020453	Dieldrin	60-57-1 OR Dieldrin				995	23	95	328	108	254	78	0
DTXSID7032638	Pyraclostrobin	175013-18-0 OR Pyraclostrobin OR pyrachlostrobin				146	10	17	5	9	15	21	0
DTXSID8024151	Imazalil	35554-44-0 OR Imazalil OR enilconazole				166	4	24	28	12	11	11	3
	Linalool	Linalool				1093	35	46	99	6	174	30	65
	Styrene	Styrene[majr]				366	444	206	496	60	200	56	4
	TPHP	Triphenylphosphate OR "triphenyl phosphate"				96	5	13	30	31	38	38	3
	PERC/TCE	Trichloroethylene[majr]				109	51	117	633	79	212	61	10
	Dronabinol/THC	Dronabinol				513	46	64	504	144	2772	64	3

Figure F4. New Landscape view

The CuratedLists sheet is a place for you to hold a large list of chemicals. (Figure F4) Let's say you are interested in a couple hundred chemicals and hoping to find out something about them and you've got valuable information to go with them … molecular weight or where you found it or whatever. On this sheet you can have the chemical identifier, preferred name, the query (which could be the same as preferred name of expanded with synonyms) and then the columns from D on are anything you want.

When you are ready to query them, select the row and click on the Send to Landscape.

	A	B	C	D	E
1	Curated chemicals		Send to Landscape	*These are sample records. Feel free to delete rows after row 3.*	
2	Copy these columns (A-C) to Landscape - Make sure column A on the Landscape sheet is not hidden			Some other columns of interest to customize	
3	DSSToxID	Preferred Name or team name	Chemical query	CAS	something else …
4	DTXSID6020430	1,2-Dichlorobenzene	1,2-Dichlorobenzene OR 95-50-1 OR o-Dichlorobenzene	*95-50-1*	
5	DTXSID1020431	1,4-Dichlorobenzene	1,4-Dichlorobenzene OR 106-46-7 OR p-Dichlorobenzene	*106-46-7*	
6	DTXSID6020438	1,2-Dichloroethane	1,2-Dichloroethane OR 107-06-2	*107-06-2*	
7	DTXSID7024031	(E)-1,2-Dichloroethylene	"(E)-1,2-Dichloroethylene" OR 156-60-5 OR trans-1,2-Dichloroethylene	*156-60-5*	
8	DTXSID5021380	1,1,2-Trichloroethane	1,1,2-Trichloroethane OR 79-00-5	*79-00-5*	
9	DTXSID0020448	1,2-Dichloropropane	1,2-Dichloropropane OR 78-87-5	*78-87-5*	
10	DTXSID1020437	1,1-Dichloroethane	1,1-Dichloroethane OR 75-34-3	*75-34-3*	
11	DTXSID3020415	1,2-Dibromoethane	1,2-Dibromoethane OR 106-93-4 OR Ethylene dibromide	*106-93-4*	
12	DTXSID3020203	1,3-Butadiene	1,3-Butadiene OR 106-99-0	*106-99-0*	
13	DTXSID8027373	Galaxolide	Galaxolide OR 1222-05-5 OR 1,3,4,6,7,8-Hexahydro-4,6,6,7,8,8-hexame	*1222-05-5*	
14	DTXSID1026081	3,3',5,5'-Tetrabromobisphenol A	3,3',5,5'-Tetrabromobisphenol A OR 79-94-7 OR 4,4'-(1-Methylethylide	*79-94-7*	
15	DTXSID5021411	Tris(2-chloroethyl) phosphate	Tris(2-chloroethyl) phosphate OR 115-96-8	*115-96-8*	
16	DTXSID1021952	Triphenyl phosphate	Triphenyl phosphate OR 115-86-6 OR Triphenylphosphate	*115-86-6*	
17	DTXSID7020637	Formaldehyde	Formaldehyde OR 50-00-0	*50-00-0*	
18	DTXSID2021159	Phthalic anhydride	Phthalic anhydride OR 85-44-9	*85-44-9*	
19	DTXSID2021781	1,2-Benzenedicarboxylic acid, dibutyl est	1,2-Benzenedicarboxylic acid, dibutyl ester OR 84-74-2 OR Dibutyl pht	*84-74-2*	
20	DTXSID3020205	Benzyl butyl phthalate	Benzyl butyl phthalate OR 85-68-7 OR 1,2-Benzene- dicarboxylic acid,	*85-68-7*	
21	DTXSID5020607	Di(2-ethylhexyl) phthalate	Di(2-ethylhexyl) phthalate OR 117-81-7 OR Di-ethylhexyl phthalate OI	*117-81-7*	
22	DTXSID9022522	Diisobutyl phthalate	Diisobutyl phthalate OR 84-69-5 OR Di-isobutyl phthalate	*84-69-5*	
23	DTXSID5025021	Dicyclohexyl phthalate	Dicyclohexyl phthalate OR 84-61-7	*84-61-7*	
24	DTXSID0028666	1,2-Benzenedicarboxylic acid, di-C9-11-br	1,2-Benzenedicarboxylic acid, di-C9-11-branched alkyl esters, C10-rich	*68515-49-1*	
25	DTXSID5028665	DINP branched	DINP branched OR 68515-48-0 OR Diisononyl phthalate	*68515-48-0*	
26	DTXSID7027205	Octamethylcyclotetrasiloxane	Octamethylcyclotetrasiloxane OR 556-67-2 OR Octamethylcyclotetra-s	*556-67-2*	
27	DTXSID4023888	Asbestos	Asbestos OR 1332-21-4	*1332-21-4*	
28	DTXSID4020533	1,4-Dioxane	1,4-Dioxane OR 123-91-1	*123-91-1*	
29					

Figure F5. CuratedLists sheet – a place to keep larger more complex lists.

G. TermMap Sheet

Term mapping is a new functionality that is motivated by the following questions that some users have asked.

- There is a set of terms I always sift on to find articles relevant to my work. Can I somehow sift on them all at once?
- I am doing research on ocean warming and its effect on fish. I have a list of over 100 fish species I am interested in. Can I find the citations quickly that have any of these fish? And because my species are in a family ontology, can I sift on the family name, too?
- Can I map genes of interest to a specialized ontology?
- How do I find the articles most about a topic like Fate and Transport?

The Term Map has built in examples. Let's walk through one of them here.

Let's say that on the Main sheet I've just run my fipronil query and have over 1000 citations. On the TermMap sheet I click on More. A form is displayed that looks like the one in Figure G1. Click on the Rodent Studies button. (Be sure to check out the other examples at some time – they'll give you an idea of the variety of ways you can use this new feature.)

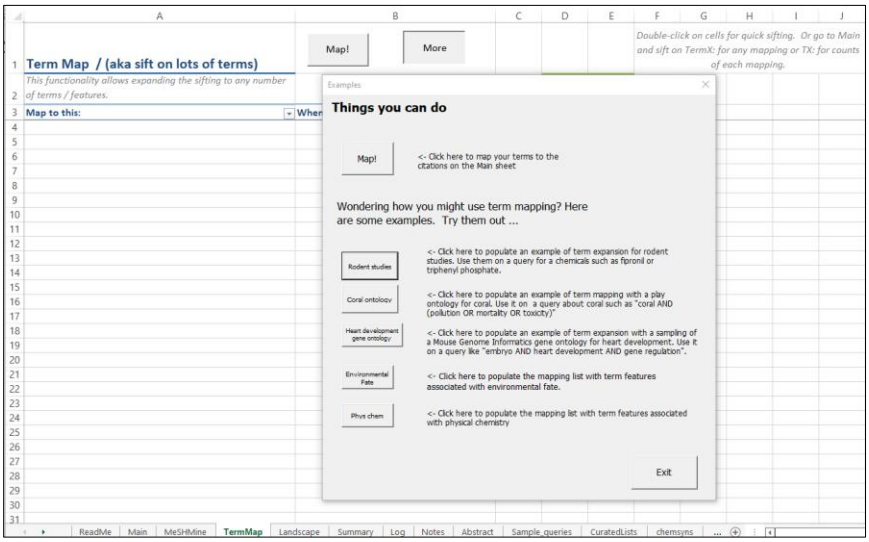


Figure G1. TermMap sheet More button actions

The TermMap sheet is then populated with terms and partial terms … let's call them features that are found in the citations of articles about rodent studies. You can edit this list – add or delete, just follow the pattern of general term and specific term.

Next, click on the Map! Button. After a few seconds you will see in numbers appear in Column C. (Figure G2.) These are counts of the number of citations from Main that contain the term in Column B. The feature Wistar, for example, is in 29 of the citations on the Main sheet.

	A	B	C	D	E
1	**Term Map / (aka sift on lots of terms)**	Map! More			
2	*This functionality allows expanding the sifting to any number of terms / features.*			Boost	
3	**Map to this:**	**When you see this:**	**Art Ct**	**by term**	**by qty**
4	Rodent	control	755		
5	Rodent	dose	386		
6	Rodent	gavage	17		
7	Rodent	mg	267		
8	Rodent	ppm	52		
9	Rodent	OECD	4		
10	Rodent	EPA	6		
11	Rodent	NTP	0		
12	Rodent	guideline	18		
13	Rodent	mice	46		
14	Rodent	mice	50		
15	Rodent	mouse	20		
16	Rodent	B6C3F1	0		
17	Rodent	rats	89		
18	Rodent	rats	97		
19	Rodent	rabbit	10		
20	Rodent	Sprague-Dawley	10		
21	Rodent	Sprague Dawley	0		
22	Rodent	fischer 344	0		
23	Rodent	Fischer-344	0		
24	Rodent	F-344	0		
25	Rodent	F344	0		
26	Rodent	Wistar	29		
27	Rodent	Long-Evans	6		
28	Rodent	inbred	7		
29	Rodent	BALB	1		

Figure G2. TermMap sheet after Map! Action has counted occurrences of the terms.

The Abstract Sifter did more than just count the articles where it found the terms. The Sifter changed the abstract for each citation in which at least one of the terms was found. (Of course, the abstract is changed only in the Sifter, not at PubMed. Rerunning the query will refresh the abstracts back to their original state.)

Here's how the abstract is modified: if any of the column B terms was found in the title and abstract, then "TERMX:" was appended to the abstract. For each column B term found, "TX: " and the column A value plus the Column B value were stuck on the end of the abstract.

To see an example that will make this clearer … see Figure G3 below of an abstract on the Abstract Sheet. At the end you see "TERMX: |TX:Rodent:control |TX:Rodent:dose |TX:Rodent:mg |TX:Rodent: rats |TX:Rodent:rats |TX:Rodent:Wistar". That's what the term mapping feature added.

	A	B
1	**Abstract with highlights**	<-- Main < > Add Note See existing notes Like this?
2	**PMID:**	14643964
4	**Title:**	Reproductive adverse effects of fipronil in Wistar rats.
	Title and Abstract:	Reproductive adverse effects of fipronil in **Wistar** **rats**. ABSTRACT: The purpose of the present study was to investigate possible reproductive adverse effects of fipronil (Frontline TopSpot) in female **Wistar** **rats**. The pesticide was topically applied to **rats** (single **dose**) at different concentrations (70, 140 and 280 **mg**/kg) and hormonal analysis, estrous cycle, and pregnancy and outcome data were determined. Treatment with fipronil altered cyclicity of female **rats** lengthening the estrous cycle (days) after a single topic administration of 70 **mg**/kg (9.7+/-1.18) or 280 **mg**/kg (14.5+/-1.45) when compared to **control** (4.8+/-0.17). In the mating study fipronil reduced the pregnancy index (67%) in the highest **dose** group (280 **mg**/kg). Plasma progesterone and estradiol levels, obtained in different periods after treatment with fipronil (70 **mg**/kg), were significantly different 96 h after treatment, when compared to **control**s. In summary, the results of the present study indicate that fipronil may alter the normal functioning of the endocrine system and cause adverse reproductive effects in female **rats**. KEYWORDS: MeSH Chemical: Insecticides|Pyrazoles|Progesterone|Estradiol|fipronil MeSH: | Animals | Estradiol/blood | Estrus/drug effects | Female | Insecticides/majr/toxicity | Pregnancy | Pregnancy Outcome | Progesterone/blood | Pyrazoles/majr/toxicity | **Rats** | **Rats**, **Wistar** | Reproduction/majr/drug effects TERMX: |TX:Rodent:**control** |TX:Rodent:**dose** |TX:Rodent:**mg** |TX:Rodent: **rats** |TX:Rodent:**rats** |TX:Rodent:**Wistar**

Figure G3. How term mapping changes the abstract. See the circled area.

In the Abstract Sifter, if text exists in the abstract column that means it can be sifted on. There are three new ways to sift if you map terms.

1. On the Main sheet, in cell B3, C3, or D3 - Sift on *TERMX:* … this will show you any citation that has a least one of the column B terms.
2. On the Main sheet, cell B3, C3, or D3 - Sift on *TX:* … this will count the number of terms from column B found in the citation.
3. Sift on a value in column A or column B. To do this quickly, stay on the TermMap sheet and double-click on a term in either A or B and see what happens. (What does happen? The Sifter copies the term to the Main sheet cell B3, lets the Main sheet sift the results, and then sorts by B3. It happens fast, but it is very simple.)

Suppose you want to find the article on the Main sheet that is most about rodents. In our definition that means the one that has the most features from our rodent feature (term) lists. Double-click on the rodent on the TermMap sheet. You'll be taken to the Main sheet, the sifter cell filled in and rows sorted (Figure G4). The records on Main are now sorted by the number of features they have connecting them to the Rodent feature/term set.

	A	B	C	D	E	F	G
1	Abstract Sifter		Query PubMed				*PubMed query run: fipronil*
2	Version 7	Your sifter terms and frequency counts					
3	PMID	TX:Rodent	rats	mg	Score	Pub Yr	Title
4	23962444	9	6		15	2013	The nuclear receptors pregnane X receptor and constitutive androstane receptor contr
5	27371222	9	7	5	21	2017	Fipronil-induced genotoxicity and DNA damage in vivo: Protective effect of vitamin E.
6	12442503	8	5	5	18	2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.
7	31873888	7	7	4	18	2020	Synergistic antioxidant effects of resveratrol and curcumin against fipronil-triggered ox
8	29854616	7	4	3	14	2018	Assessment of fipronil toxicity on wistar rats: A hepatotoxic perspective.
9	33201557	7	8		15	2021	Efficacy of a combined insecticide-rodenticide product on ectoparasite and commensa
10	31428841	7	7	2	16	2019	Blood pharmacokinetic of 17 common pesticides in mixture following a single oral expo
11	29368479	7		3	10	2016	[Genotoxic effects of pesticide fipronil in somatic and generative cells of mice].
12	14643964	6	9	6	21	2004	Reproductive adverse effects of fipronil in Wistar rats.
13	32772290	6	9	3	18	2020	Ginseng attenuates fipronil-induced hepatorenal toxicity via its antioxidant, anti-apopt
14	30290348	6	7	4	17	2019	Hepatorenal protective effects of taurine and N-acetylcysteine against fipronil-induced
15	26409903	6	7	2	15	2015	Prenatal exposure to fipronil disturbs maternal aggressive behavior in rats.
16	34475715	6	7	9	22	2021	Synergistic effects of sitagliptin and losartan against fipronil-induced hepatotoxicity in r
17	31881178	6	5	2	13	2020	Transcriptomic modifications of the thyroid gland upon exposure to phytosanitary-grad
18	27166213	6	8	2	16	2017	Perinatal exposure to insecticide fipronil: effects on the reproductive system in male ra
19	27074097	6	6	4	16	2016	In utero and lactational exposure to fipronil in female rats: Pregnancy outcomes and se
20	26142839	6	7	2	15	2015	Fipronil induces CYP isoforms in rats.
21	34493330	6			6	2021	Efficacy of low-dose fipronil bait against blacklegged tick (Ixodes scapularis) larvae fee
22	32784929	6	5	2	13	2020	Chronic Administration of Fipronil Heterogeneously Alters the Neurochemistry of Mon
23	31152858	6	5	4	15	2019	Garlic and allopurinol attenuate hepatic apoptosis induced by fipronil in male albino rat
24	26960908	6	5	2	13	2016	DBS-platform for biomonitoring and toxicokinetics of toxicants: proof of concept using

Figure G4. After double-clicking on Rodent on the TermMap sheet.

H. MeSHMine sheet

MeSH mining is a feature to help answer the needs of researchers who ask questions like these:

- I ran a query on Histamine H2 antagonism. Can I get a list of all the chemicals annotated in those PubMed records?
- I ran a query on unfolded protein response and chemical toxicity. I'd like to find some chemicals that cause this response. Is there a fast way to do this?
- I ran a query on retinal disease. Can I get a list of genes and proteins annotated in my resulting article set on Main?
- I ran a query on biosolids. I want to see a list of chemicals found in biosolids.

As the above list shows, step one in using this new feature is to run a query to the Main sheet. Let's walk through an example.

Let's say we ran the query (Anthozoa OR coral) AND (chemical toxicity OR drug effects).

Running this query brings back 2,459 records to the Main sheet. Next, click on the button More things on the top of the Main sheet. You'll see this form in Figure H1. Click on Extract MeSH Substances. The next form asks whether you want to extract all annotated chemicals or just the major topics. (If you don't know what this means, read about annotation at the PubMed web site.)

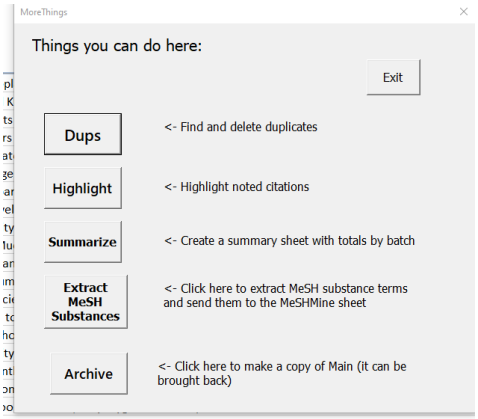


Figure H1. Click on Extract MeSH Substances.

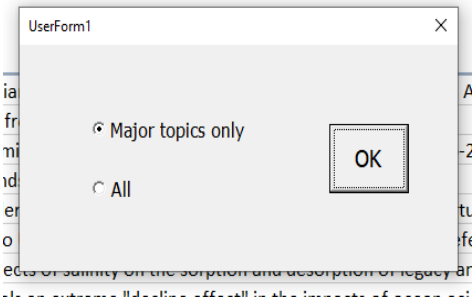
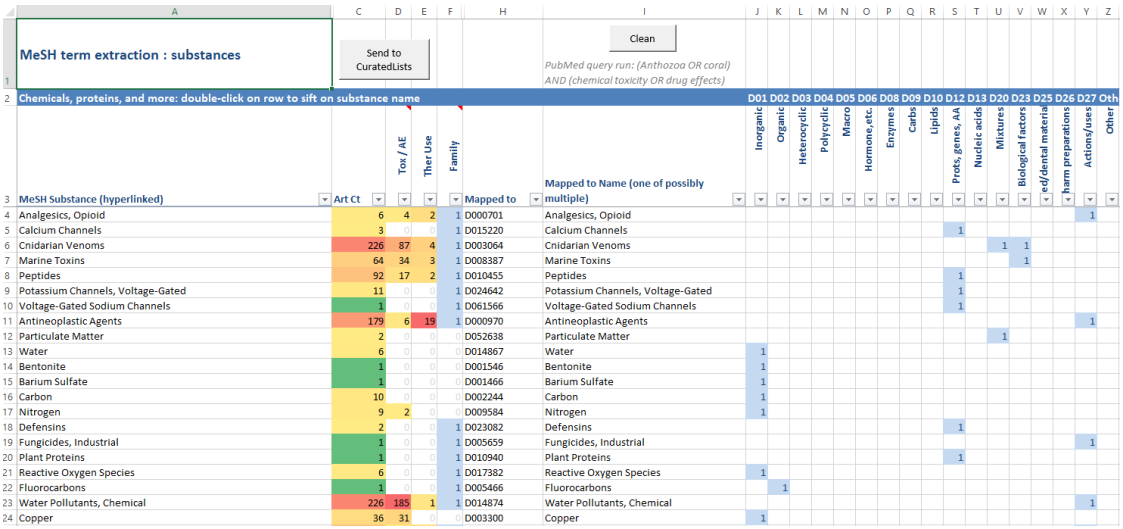


Figure H2. Try both options for fun.

In a few minutes or seconds (depending on the size of the collection on Main and your internet speed) you will be taken to the MeSHMine sheet and a total displayed.

Let's see what's on the MeSHMine sheet (Figure H3).



MeSH term extraction : substances

Send to CuratedLists

Clean

PubMed query run: (Anthozoa OR coral) AND (chemical toxicity OR drug effects)

Chemicals, proteins, and more: double-click on row to sift on substance name

MeSH Substance (hyperlinked)	Art Ct	Tox / AE	Ther Use	Family	Mapped to	Mapped to Name (one of possibly multiple)	D01 Inorganic	D02 Organic	D03 Heterocyclic	D04 Polycyclic	D05 Macro	D06 Hormone,etc.	D08 Enzymes	D09 Carbs	D10 Lipids	D12 Prots, genes, AA	D13 Nucleic acids	D20 Mixtures	D23 Biological factors	D25 ed/dental materials	D26 harm preparations	D27 Actions/uses	Oth Other
Analgesics, Opioid	6	4	2	1	D000701	Analgesics, Opioid																1	
Calcium Channels	3	0	0	1	D015220	Calcium Channels										1							
Cnidarian Venoms	226	87	4	1	D003064	Cnidarian Venoms												1	1				
Marine Toxins	64	34	3	1	D008387	Marine Toxins													1				
Peptides	92	17	2	1	D010455	Peptides										1							
Potassium Channels, Voltage-Gated	11	0	0	1	D024642	Potassium Channels, Voltage-Gated										1							
Voltage-Gated Sodium Channels	1	0	0	1	D061566	Voltage-Gated Sodium Channels										1							
Antineoplastic Agents	179	6	19	1	D000970	Antineoplastic Agents																1	
Particulate Matter	2	0	0	0	D052638	Particulate Matter												1					
Water	6	0	0	0	D014867	Water	1																
Bentonite	1	0	0	0	D001546	Bentonite	1																
Barium Sulfate	1	0	0	0	D001466	Barium Sulfate	1																
Carbon	10	0	0	0	D002244	Carbon	1																
Nitrogen	9	2	0	0	D009584	Nitrogen	1																
Defensins	2	0	0	1	D023082	Defensins										1							
Fungicides, Industrial	1	0	0	1	D005659	Fungicides, Industrial																1	
Plant Proteins	1	0	0	1	D010940	Plant Proteins										1							
Reactive Oxygen Species	6	0	0	1	D017382	Reactive Oxygen Species	1																
Fluorocarbons	1	0	0	1	D005466	Fluorocarbons		1															
Water Pollutants, Chemical	226	185	1	1	D014874	Water Pollutants, Chemical																1	
Copper	36	31	0	0	D003300	Copper	1																

Figure H3. MeSHMine sheet.

Column A has the substance term. Try sorting on it. It can be a MeSH heading or MeSH supplemental concept (again, go to the PubMed help to find out the difference.) Each name is hyperlinked to the MeSH browser. You can click on the name to go to the browser to find out more.

(A quick aside about MeSH supplemental concepts versus MeSH headings. Most chemicals are supplemental concepts and they are mapped to MeSH heading terms. Part of the process in creating this sheet is to take each concept and behind the scenes, map it to one of its MeSH heading terms. Once a concept is mapped, it inherits tree structure and annotated subheadings from its parent.)

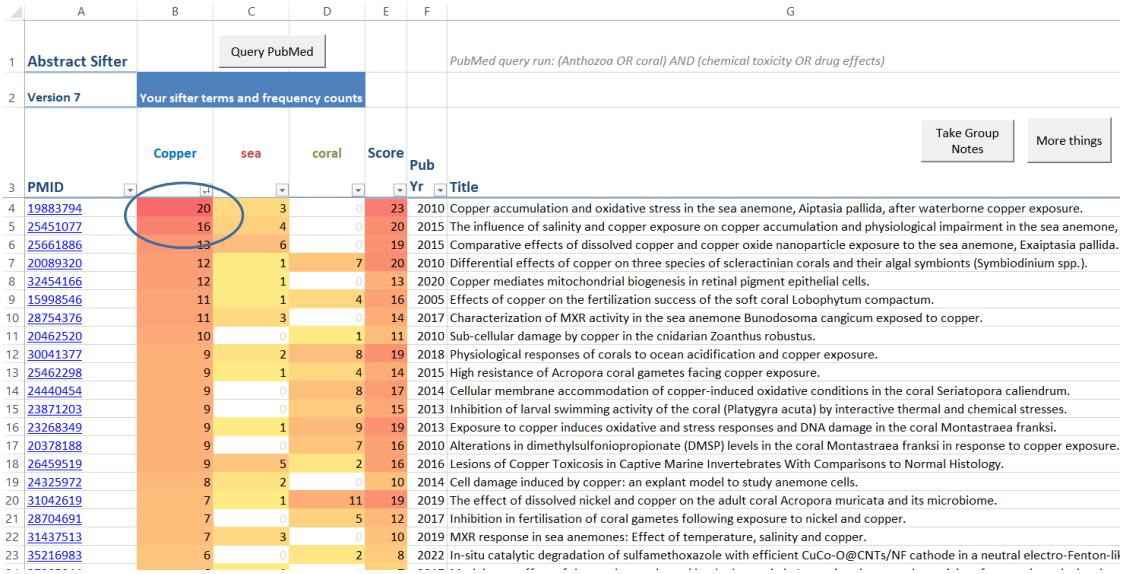
The ArtCt column has the total number of articles on the Main sheet in which the MeSH terms was annotated. (Note – this is not the same as occurrences in the text of the abstract.) The Tox/AE column counts the number of articles on Main in which the substance appeared with the co-annotation toxicity or adverse effects or poisoning. The next column titled Ther Use counts the articles in which the substance is co-annotated with the subheading therapeutic use. Try sorting on these columns.

The column entitled Family contains a flag: 1 = family name ; 0 = not a family name. If you are interested in individual chemicals only, you can sort by this column and delete rows. Or you can click on the Clean button above and remove them.

The rest of the columns contain some more information about the chemical and its place in the MeSH tree structure. If you are not familiar with the MeSH tree structure, go here and browse: https://meshb.nlm.nih.gov/treeView These columns can be used to sort the list. For instance, if you are only interested in genes, proteins, peptides, amino acids, then sort (descending) on column S.

Naturally, you'll want to go back to the Main sheet and read more about the substances you find. To do this quickly, double-click anywhere in the row (other than column A). This action takes you to the Main sheet and fills in the first sifter cell with the substance name and sorts.

Double-clicking on the Copper row gives me:



	A	B	C	D	E	F	G
1	Abstract Sifter		Query PubMed				*PubMed query run: (Anthozoa OR coral) AND (chemical toxicity OR drug effects)*
2	Version 7	Your sifter terms and frequency counts					
3	PMID	Copper	sea	coral	Score	Pub Yr	Title
4	19883794	20	3	0	23	2010	Copper accumulation and oxidative stress in the sea anemone, Aiptasia pallida, after waterborne copper exposure.
5	25451077	16	4	0	20	2015	The influence of salinity and copper exposure on copper accumulation and physiological impairment in the sea anemone,
6	25661886	13	6	0	19	2015	Comparative effects of dissolved copper and copper oxide nanoparticle exposure to the sea anemone, Exaiptasia pallida.
7	20089320	12	1	7	20	2010	Differential effects of copper on three species of scleractinian corals and their algal symbionts (Symbiodinium spp.).
8	32454166	12	1	0	13	2020	Copper mediates mitochondrial biogenesis in retinal pigment epithelial cells.
9	15998546	11	1	4	16	2005	Effects of copper on the fertilization success of the soft coral Lobophytum compactum.
10	28754376	11	3	0	14	2017	Characterization of MXR activity in the sea anemone Bunodosoma cangicum exposed to copper.
11	20462520	10	0	1	11	2010	Sub-cellular damage by copper in the cnidarian Zoanthus robustus.
12	30041377	9	2	8	19	2018	Physiological responses of corals to ocean acidification and copper exposure.
13	25462298	9	1	4	14	2015	High resistance of Acropora coral gametes facing copper exposure.
14	24440454	9	0	8	17	2014	Cellular membrane accommodation of copper-induced oxidative conditions in the coral Seriatopora caliendrum.
15	23871203	9	0	6	15	2013	Inhibition of larval swimming activity of the coral (Platygyra acuta) by interactive thermal and chemical stresses.
16	23268349	9	1	9	19	2013	Exposure to copper induces oxidative and stress responses and DNA damage in the coral Montastraea franksi.
17	20378188	9	0	7	16	2010	Alterations in dimethylsulfoniopropionate (DMSP) levels in the coral Montastraea franksi in response to copper exposure.
18	26459519	9	5	2	16	2016	Lesions of Copper Toxicosis in Captive Marine Invertebrates With Comparisons to Normal Histology.
19	24325972	8	2	0	10	2014	Cell damage induced by copper: an explant model to study anemone cells.
20	31042619	7	1	11	19	2019	The effect of dissolved nickel and copper on the adult coral Acropora muricata and its microbiome.
21	28704691	7	0	5	12	2017	Inhibition in fertilisation of coral gametes following exposure to nickel and copper.
22	31437513	7	3	0	10	2019	MXR response in sea anemones: Effect of temperature, salinity and copper.
23	35216983	6	0	2	8	2022	In-situ catalytic degradation of sulfamethoxazole with efficient CuCo-O@CNTs/NF cathode in a neutral electro-Fenton-li

Take Group Notes | More things

Figure H4. Double-clicking on Copper on the MeSHMine sheet brings you here.

To get back to the MeSHMine sheet, just select it below.

When you've found the substances you want to work with more, copy the names and paste them onto the Landscape sheet Column C. Or you can select rows on MeSHMine and click on the button Send to Curate. This is a sheet used to keep a list of things of interest. It holds the query language and any other information you want to keep track of. From the CuratedLists sheet, select rows and Send to Landscape for in-depth examination.

I. Exporting to other applications from the Notes sheet

The Abstract Sifter allows the user to export articles from the Notes sheet to outside applications. On the Notes sheet there is a button labeled Export. By clicking on this button, the form in Figure 26 appears. The first set of radio buttons allows the user to select what data is exported and how the records should be formatted. Next the user can choose to export all entries / rows on the Notes sheet or just selected rows. If the end-user selects PMIDs to be exported, the PMIDs will be formatted in the box. In this case, clicking on Next Step will copy the formatted PMIDs to the clipboard, ready to be pasted to the next application. They should then be pasted into the destination. In the case of PubMed, they should be pasted into the query box in PubMed Entrez. (Figure I1) From PubMed, the citations can be downloaded in a variety of formats, including a format that can be imported into citation management software (Figures I2 and I3).

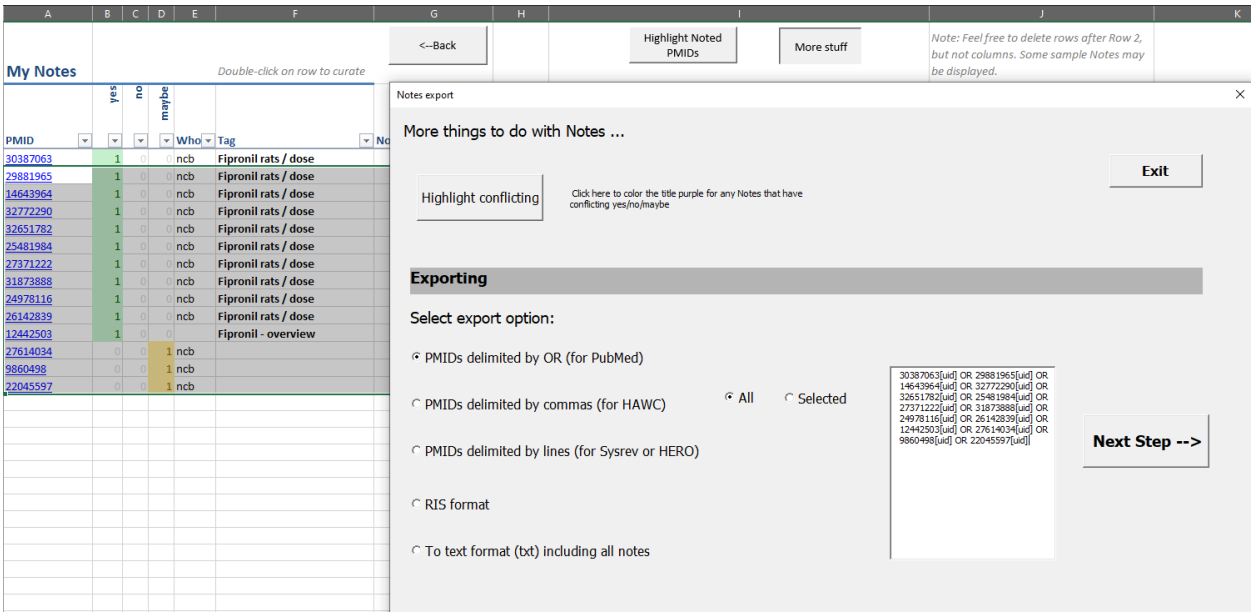


Figure I1. Form that appears after clicking on Export button

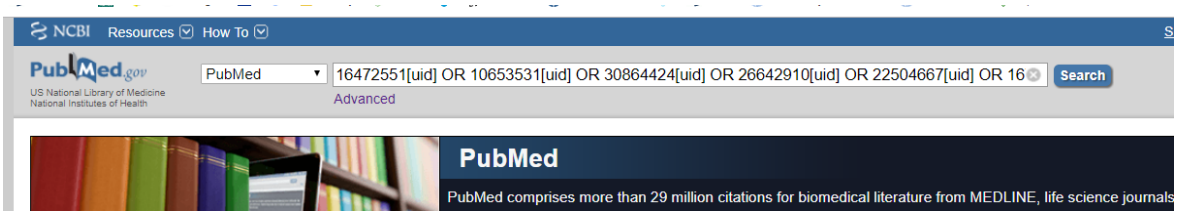


Figure 24. Paste PMIDs in the query box to retrieve the records.

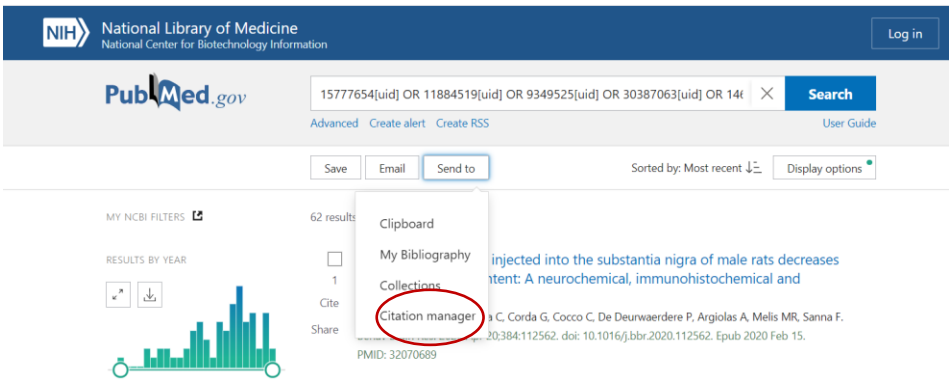


Figure I2 In PubMed, click on Send to

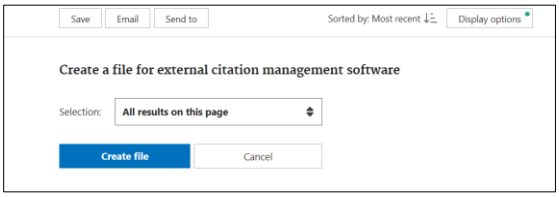


Figure I3. Dialog box for exporting to citation manager

When the user clicks on the Create File button, a file is created and downloaded in nbib format and can be imported into most common reference manager software.

If the end-user specifies RIS format or text format, files will be created. The user will specify the location and name of the files. The text file option includes the abstract, tags, and notes.

J. Helpful Tips and Guidelines

Tip 1 – Checking quality of results

The Landscape sheet is a great way to explore a set of chemicals, but some chemical names are long, complex, and a challenge to PubMed. If you copy and paste a chemical name from another source, make sure it does not have any special characters. Non printing escape characters make the web service calls give unexpected results, but the PubMed web site knows how to ignore them. Also, try your chemical names with quotation marks and without them to see which works better. Sometimes quotation marks are essential.

For other types of errors or strange results, it's a good idea to check it in PubMed. You can take any query generated by the Abstract Sifter and copy and paste it into PubMed using Ctl-C to copy and Ctl-V to paste. For example, the query in the box shown in Figure J1 is selected and copied (with Ctl-C). Then in PubMed the query is pasted into the query line at the top as shown in Figure J2. On the right side of the page is a box entitled Search Details. Click on the See More … link to expand this box. Figure J3 shows the information provided by PubMed about how it expands the query. If you need to learn more about PubMed queries, click on Help on the PubMed home page.

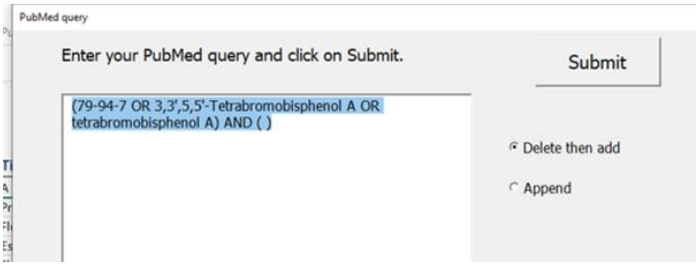


Figure J1. Select and Ctl-C to copy

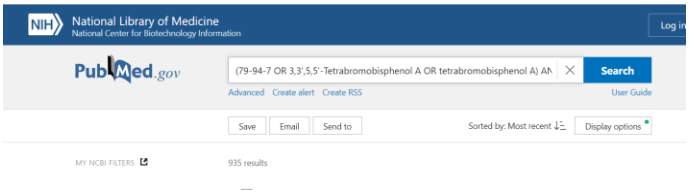


Figure J2. Ctl-V to paste in PubMed then search

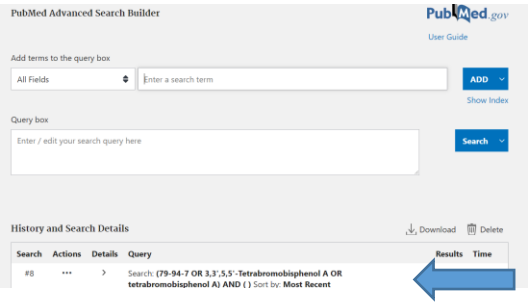


Figure J3. See what PubMed does to expand your search

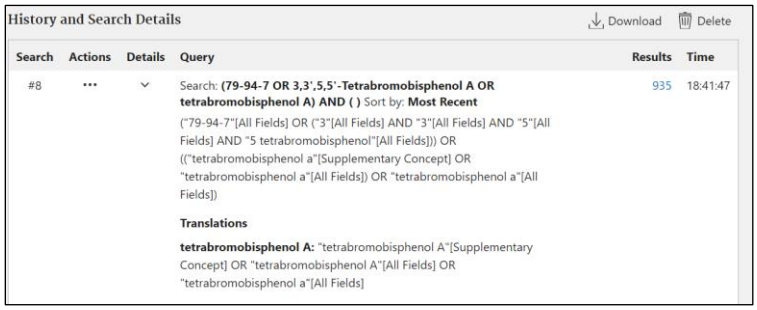


Figure J4. PubMed query breakdown and expansion

Note that after PubMed rolled out its new version in 2020, not just the looks of the site changed. They redid the search algorithms. As of this writing, the web service had not changed to use the new search methods – that means for the first time we are noticing small differences in the returned record counts.

Tip 2 – Sifting the chemical literature

It can be very helpful in chemical research to include the chemical name in the sifting process. This is because a chemical can be mentioned in an abstract even in cases where the article is not really about the chemical and will be retrieved in the PubMed query (depending on how the query is worded). Counting the occurrences of the chemical name in the abstract through the sifting process can help the user discriminate between articles mentioning a chemical or those that are actually about the chemical.

Tip 3 – Cleanup and customization

The Abstract Sifter can be cleaned up by deleting rows and columns from previous work, but the Abstract Sifter programming requires certain columns and rows to be in certain places. To learn how to clean up your sifter without disrupting the behind-the-scenes coding, consult the table below.

Sheet name	**Advice for cleaning**
Main	Do not add columns. Rows will be added and deleted by the Sifter.
Abstract	Do not add or delete rows or columns. The Sifter software updates this sheet.

Notes	Delete any unwanted rows after Row 2. Do not add or delete columns.
Log	Delete any unwanted rows after Row 2. Do not delete columns. Add columns after G if desired.
Sample_Queries	Delete unwanted rows after Row 3. Modify and add rows as desired, following the pattern of current rows. (That is, keep the heading in column B and the query text in column C.) *Hint: use this sheet to keep queries important to your organization.*
Pathway_queries	Delete unwanted rows after Row 3. Modify and add rows as desired, following the pattern of current rows.
Landscape	Delete or modify rows after Row 4 and columns after Column D.
CuratedLists	Delete or modify rows after Row 3 and columns after Column C.
TermMap	Delete or modify rows after Row 3.
MeSHMine	Abstract Sifter deletes from this sheet before writing new data. Feel free to make a copy of it.

Keep in mind that the Abstract Sifter is an Excel file. You can rename it, mail it, and of course, if you want to keep your Log, Notes, and Landscape entries, you should save it. The Sample_queries sheet provides an opportunity for you and your organization to start collecting and organizing queries that you have found useful. As mentioned earlier, feel free to hide any sheets you don't use to keep your workplace streamlined.

Tip 4 – Collaborative literature review tips

Sometimes more than one person will want to work together on evaluating a set of articles. The Abstract Sifter has some features to make this easier. So, let's say Mary and Joe each retrieve, sift, and take notes on their own Abstract Sifter files. Mary can copy Joe's notes to her version of the Sifter and then she has both sets. (Or they can mail the Sifter back and forth …)

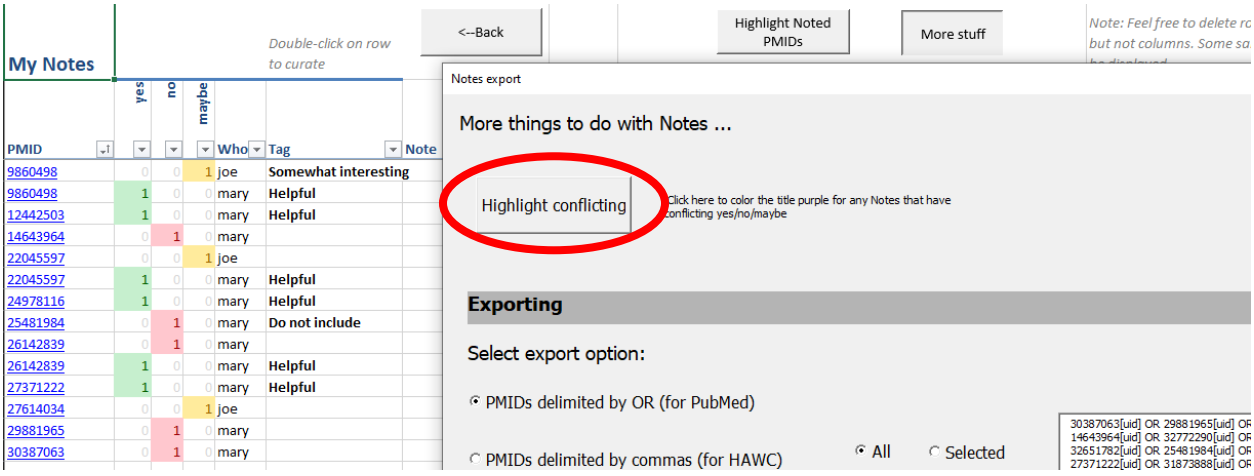
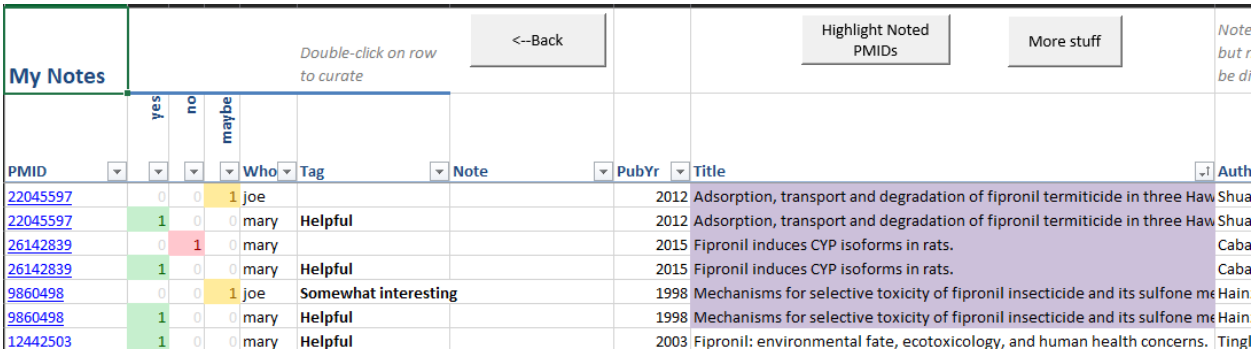


Figure J5. Notes from two reviewers combined on the Notes sheet.

To see if she and Joe disagreed on any record, she can click on More Stuff then Highlight Conflicting. The titles of Notes with different yes/no/maybe designations are colorized in purple. Resolving the conflicts and re-clicking on the button will cause the purple to disappear.



PMID	yes	no	maybe	Who	Tag	Note	PubYr	Title	Auth
22045597	0	0	1	joe			2012	Adsorption, transport and degradation of fipronil termiticide in three Haw	Shua
22045597	1	0	0	mary	Helpful		2012	Adsorption, transport and degradation of fipronil termiticide in three Haw	Shua
26142839	0	1	0	mary			2015	Fipronil induces CYP isoforms in rats.	Caba
26142839	1	0	0	mary	Helpful		2015	Fipronil induces CYP isoforms in rats.	Caba
9860498	0	0	1	joe	Somewhat interesting		1998	Mechanisms for selective toxicity of fipronil insecticide and its sulfone me	Hain
9860498	1	0	0	mary	Helpful		1998	Mechanisms for selective toxicity of fipronil insecticide and its sulfone me	Hain
12442503	1	0	0	mary	Helpful		2003	Fipronil: environmental fate, ecotoxicology, and human health concerns.	Tingl

Figure J6. Purple highlighting on title of conflicting notes.

Tip 5 – Connections to the EPA Chemicals Dashboard

The Environmental Protection Agency's Chemicals Dashboard is a great place to find chemical information to enhance your chemical search queries with synonyms and CAS numbers. Future releases of the Dashboard will offer opportunities to download a list of chemicals formatted for easy insertion into the Landscape sheet. You'll find the Chemistry Dashboard here: https://comptox.epa.gov/dashboard.

The EPA Chemicals Dashboard also contains its own (slightly different) version of the Abstract Sifter. It works on the same basic premise as the Excel version, but has some interesting differences. To see it, start with a chemical search. Let's look at the chemical fipronil by entering the name in the search box and clicking on the search icon (magnifying glass) (Figure J7).

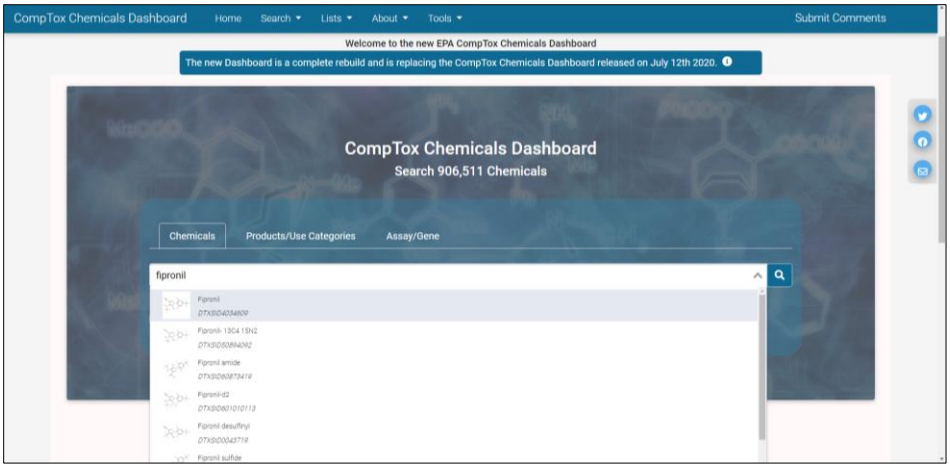


Figure J7. Searching for fipronil on the EPA's Chemistry Dashboard entry form

The main page for fipronil is displayed with the structure diagram and a selection of tabs below that lead to other information about the chemical. Click on the Literature tab as shown in Figure J8.

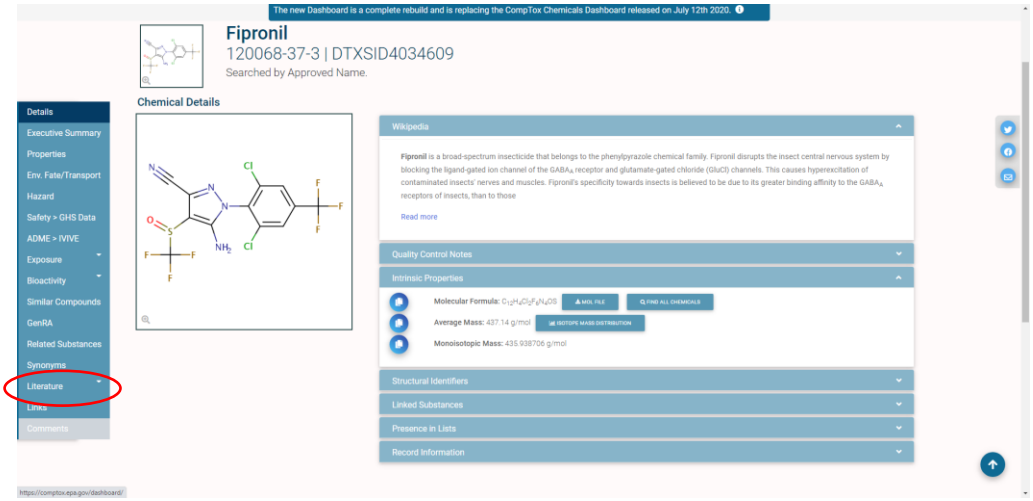


Figure J8. Select the Literature tab then on PubMed Abstract Sifter (see below).

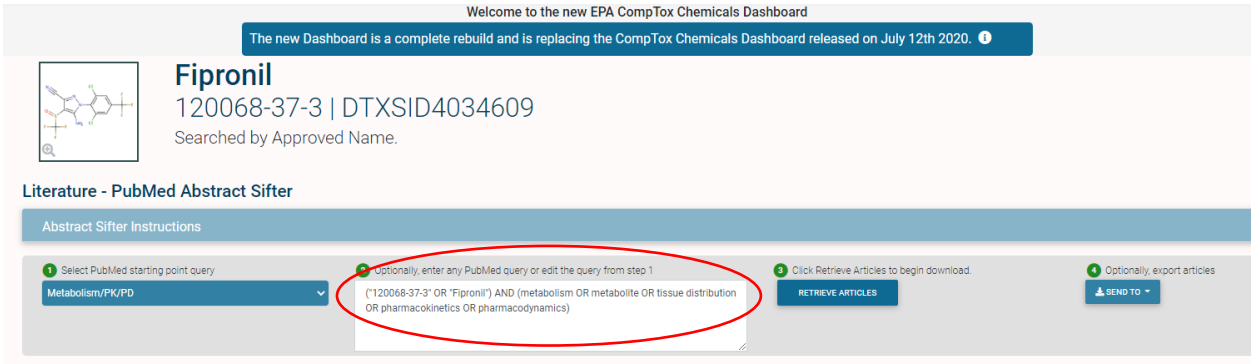
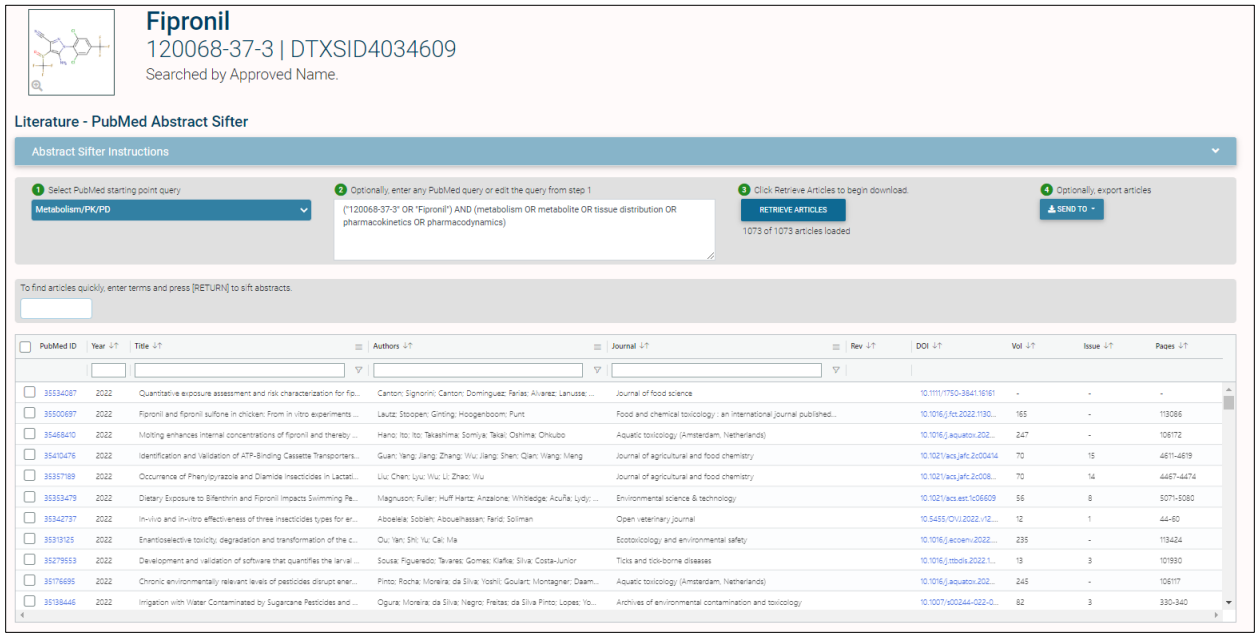


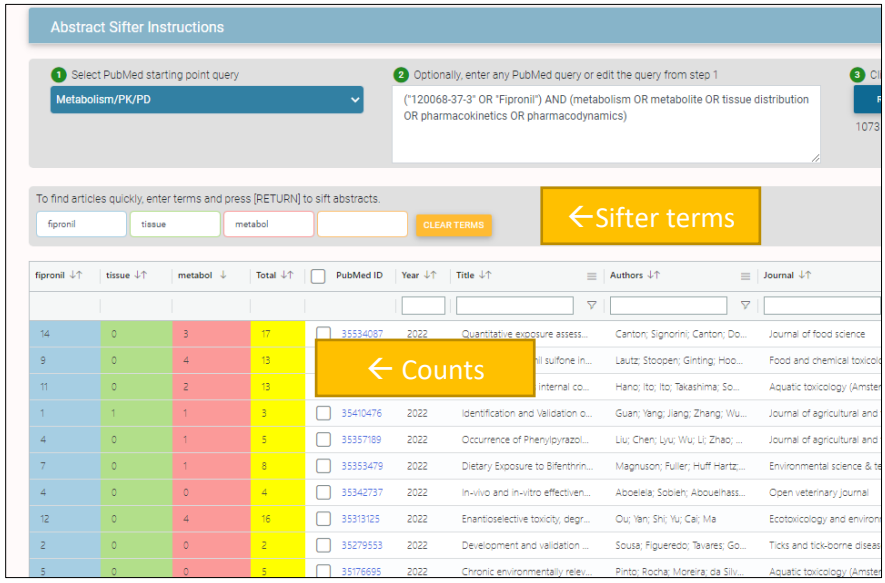
Figure J9. How to select prepared queries.

Select PubMed Abstract Sifter on the left set of buttons. The Dashboard helps you to build queries for this chemical. The chemical identifier part of the query is prepopulated on the right with name and CAS number. The subject matter part of the query is determined by selecting a topic area in the pull-down box in the center of the form. The user has several pre-composed queries to choose from. When one of them is chosen, the query is modified by appending the subject matter text. Figure J9 shows that when Metabolism/PK/PD is chosen, the text (metabolism OR metabolite OR tissue distribution OR

pharmacokinetics OR pharmacodynamics) is appended to the chemical identifiers. The query can be modified manually as well. When ready, the user clicks on Retrieve Articles. See Figure J10.



J10. Results after retrieving articles.



J11. Sifting and sorting.

Figure J10. Sifting on the EPA Chemicals Dashboard's PubMed Abstract Sifter.

After the user clicks on Retrieve Articles, the article information is retrieved from PubMed and inserted into the results table. The articles can be sifted by entering terms into the boxes shown. In the example

in Figure J10, the user has entered “fipronil”, “tissue”, and “metabol” into the sifter entry boxes. The occurrences of these terms are counted for each PubMed citation and displayed. The table can be sorted on these values. Clicking on a row tells the Dashboard Sifter to display the title and abstract below the table with the sifter terms highlighted as in Figure J11.

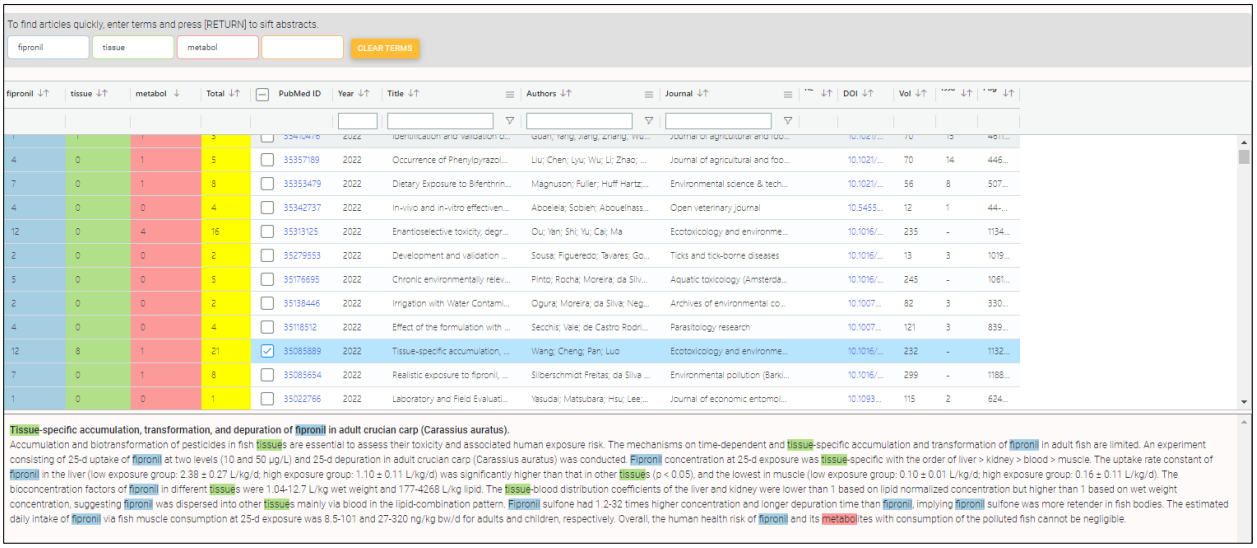


Figure J11. Clicking on a row displays the highlighted abstract below.

A check box on the left of the table provides a way to select citation rows. Selected rows can be downloaded or sent to PubMed by clicking on the pull-down box to the right.

Tip 6. Populating the Excel Abstract Sifter from the EPA CompTox Chemicals Dashboard

The Chemicals Dashboard can download chemicals in a variety of formats. One of those formats make it easy to use in the Excel Abstract Sifter. Here’s an example to get you started. On the home page of the Dashboard, click on Lists, then chemical, then pick a list. We’ll pick Algal Toxins as a sample. Click on the name, then, when the chemicals appear, click on Send to Batch Search. (OK you could have put your own list of chemicals into the Batch search … another option.)

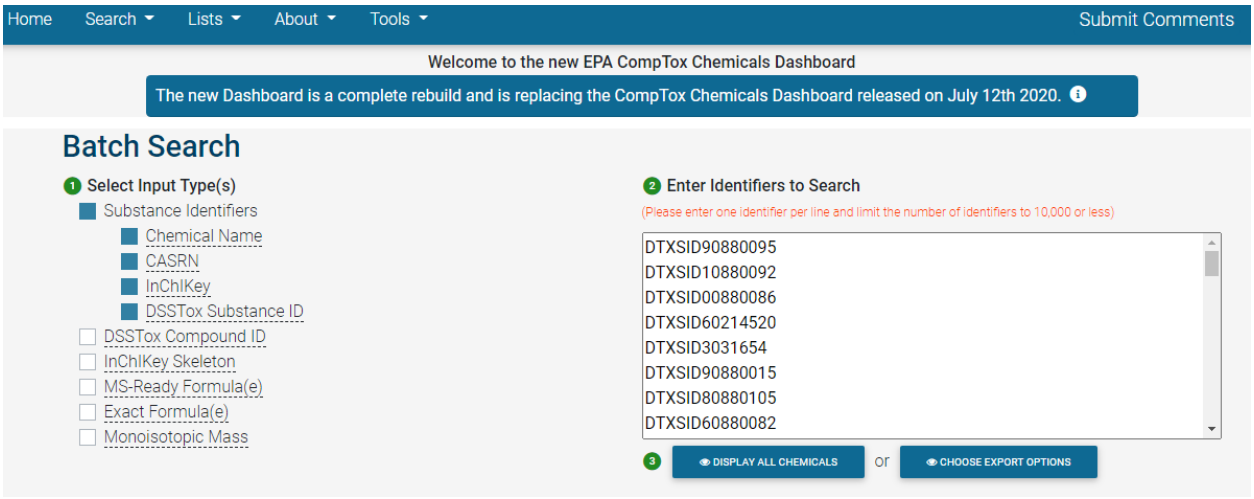


Figure J12. Send a list of chemicals to Batch Search.

Click on the following: Choose Export Options then Choose Export Format then Excel then Abstract Sifter Input file under Enhanced Data Sheets. Then Download. An Excel file will be built. Open it and go to the sheet Abstract Sifter Sheet. Copy the chemicals you want and paste them into the Abstract Sifter either on the Landscape sheet or the CuratedChemicals sheet.

Download Chemical Data, then Download as Excel, then Abstract Sifter Input File (Beta), then (finally) the Download bar. This action will download the chemicals to Excel. Open that file. It will have 2 sheets. Open the one that is called Abstract Sifter. It looks like Figure J15. On the Abstract Sifter Landscape sheet, unhide column A. This is done by clicking on the left border of Column B, then right-clicking to see the menu where you can click on *Unhide*. Paste rows from the downloaded spreadsheet onto the Landscape sheet as in Figure J16.

	A	B	C
1	DSSTOX_LINK	PREFERRED_NAME	CHEMICAL/ENTITY_QUERY
2	DTXSID2031083	Cylindrospermopsin	143545-90-8 OR Cylindrospermopsin
3	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR OR cyanoginosin LR
4	DTXSID3031656	Microcystin LA	96180-79-9 OR Microcystin LA OR cyanoginosin-LA
5	DTXSID9040974	Azaspiracid	214899-21-5 OR Azaspiracid
6	DTXSID3074313	Saxitoxin	35523-89-8 OR Saxitoxin
7	DTXSID60166611	beta-N-Methylamino-L-alanine	15920-93-1 OR beta-N-Methylamino-L-alanine
8	DTXSID70207660	Decarbamylsaxitoxin	58911-04-9 OR Decarbamylsaxitoxin
9	DTXSID60214520	Gonyautoxin V	64296-25-9 OR Gonyautoxin V
10	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid OR domoic acid
11	DTXSID90423027	palytoxin	77734-91-9 OR palytoxin
12	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a OR anatoxin I
13	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A
14	DTXSID20879997	Brevetoxin 2	79580-28-2 OR Brevetoxin 2
15	DTXSID40879999	Brevetoxin C	98225-48-0 OR Brevetoxin C
16	DTXSID40880000	Ciguatoxin 1	11050-21-8 OR Ciguatoxin 1
17	DTXSID00880001	Dinophysistoxin 1	81720-10-7 OR Dinophysistoxin 1
18	DTXSID60880002	Okadaic acid	78111-17-8 OR Okadaic acid
19	DTXSID10880012	Maitotoxin	59392-53-9 OR Maitotoxin
20	DTXSID90880015	Lyngbyatoxin-a	70497-14-2 OR Lyngbyatoxin-a
21	DTXSID10880017	Euglenophycin	1219817-69-2 OR Euglenophycin
22	DTXSID60880022	Nodularin	118399-22-7 OR Nodularin
23	DTXSID20880023	Yessotoxin	112514-54-2 OR Yessotoxin
24	DTXSID80880024	Azaspiracid 4	344422-49-7 OR Azaspiracid 4
25	DTXSID60880082	Aplysiatoxin	52659-57-1 OR Aplysiatoxin

Worksheet1 | Abstract Sifter

Figure J15. Excel view of downloaded chemicals on Abstract Sifter sheet.

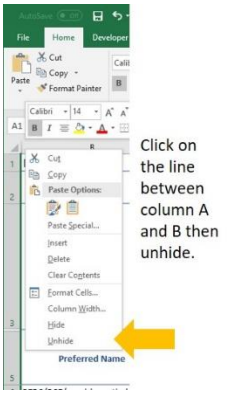


Figure J16. Unhide column A on the Landscape sheet in order to paste the DSSTox number there.

	A	B	C	E	F	G
1	Abstract Sifter	Landscape View				
2			Update Article Counts / View / hide queries / Heat Map by column / Heat Map by row			
3			Subject queries:	(dna/drug effects OR DNA Damage OR chromosome aberrations OR	neoplasms or cancer	(reproduction AND (toxicity OR abnormal OR adverse effects))
5	DSSTOX link to Dashboard	Preferred Name	Chemical / Entity query	Genetox	Cancer	ReproTox
6	DTXSID2031083	Cylindrospermopsin	143545-90-8 OR Cylindrospermopsin			
7	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR OR cyanoginosin LR			
8	DTXSID3031656	Microcystin LA	96180-79-9 OR Microcystin LA OR cyanoginosin-LA			
9	DTXSID9040974	Azaspiracid	214899-21-5 OR Azaspiracid			
10	DTXSID3074313	Saxitoxin	35523-89-8 OR Saxitoxin			
11	DTXSID60166611	beta-N-Methylamino-L-alanine	15920-93-1 OR beta-N-Methylamino-L-alanine			
12	DTXSID70207660	Decarbamylsaxitoxin	58911-04-9 OR Decarbamylsaxitoxin			
13	DTXSID60214520	Gonyautoxin V	64296-25-9 OR Gonyautoxin V			
14	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid OR domoic acid			
15	DTXSID90423027	palytoxin	77734-91-9 OR palytoxin			
16	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a OR anatoxin I			
17	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A			
18	DTXSID20879997	Brevetoxin 2	79580-28-2 OR Brevetoxin 2			
19	DTXSID40879999	Brevetoxin C	98225-48-0 OR Brevetoxin C			
20	DTXSID40880000	Ciguatoxin 1	11050-21-8 OR Ciguatoxin 1			

Figure J17. This is what the sheet will look like after unhiding Column A and pasting the chemicals downloaded from the Dashboard.

Now, enter subject matter queries, or, if you already have queries in place, select the intersecting cells and click on *Update Article Counts*. Click on one of the Heat Map buttons to make it pretty.

	A	B	C	E	F	G
1	**Abstract Sifter**	**Landscape View**				
2			Update Article Counts / View / hide queries / Heat Map by column / Heat Map by row			
3			***Subject queries:***	(dna/drug effects OR DNA Damage OR chromosome aberrations OR	neoplasms or cancer	(reproduction AND (toxicity OR abnormal OR adverse effects))
5	**DSSTOX link to Dashboard**	**Preferred Name**	**Chemical / Entity query**	**Genetox**	**Cancer**	**ReproTox**
6	DTXSID2031083	Cylindrospermopsin	143545-90-8 OR Cylindrospermopsin	48	15	8
7	DTXSID3031654	Microcystin LR	101043-37-2 OR Microcystin LR OR cyanoginosin LR	89	129	74
8	DTXSID3031656	Microcystin LA	96180-79-9 OR Microcystin LA OR cyanoginosin-LA	1	1	0
9	DTXSID9040974	Azaspiracid	214899-21-5 OR Azaspiracid	2	12	1
10	DTXSID3074313	Saxitoxin	35523-89-8 OR Saxitoxin	20	51	5
11	DTXSID60166611	beta-N-Methylamino-L-alanine	15920-93-1 OR beta-N-Methylamino-L-alanine	13	13	10
12	DTXSID70207660	Decarbamylsaxitoxin	58911-04-9 OR Decarbamylsaxitoxin	0	0	0
13	DTXSID60214520	Gonyautoxin V	64296-25-9 OR Gonyautoxin V	0	0	0
14	DTXSID20274180	L-Domoic acid	14277-97-5 OR L-Domoic acid OR domoic acid	78154	279707	12022
15	DTXSID90423027	palytoxin	77734-91-9 OR palytoxin	6	30	2
16	DTXSID50867064	Anatoxin a	64285-06-9 OR Anatoxin a OR anatoxin I	6	4	6
17	DTXSID60879996	Brevetoxin A	98112-41-5 OR Brevetoxin A	1	2	0
18	DTXSID20879997	Brevetoxin 2	79580-28-2 OR Brevetoxin 2	4	3	1
19	DTXSID40879999	Brevetoxin C	98225-48-0 OR Brevetoxin C	6	14	4
20	DTXSID40880000	Ciguatoxin 1	11050-21-8 OR Ciguatoxin 1	2	30	2

Figure J18. Downloaded chemicals and queries with subject matter queries.

Now you have an overview of your chemicals and what literature is out in PubMed for them. Take advantage of the iterative nature of the Abstract Sifter to query, sift, read, note as much as you need.

K. In-depth discussions

MeSH Mining in-depth

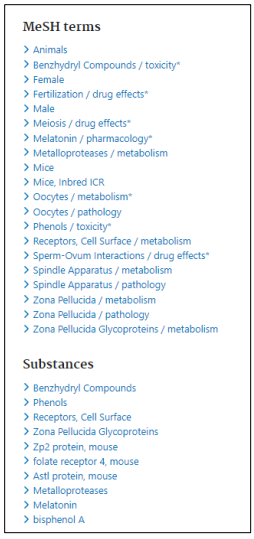
Most PubMed citations are enhanced by the addition of MeSH indexing terms. This indexing is performed by a combination of automated text-mining and manual curation. One of the entities indexed is substance name. Substances (like chemicals) discussed in the abstract will likely have the chemical name annotated, often along with a family name for that chemical. Note: publications that discuss lots of chemicals in the text (i.e. "we studied 25 pesticides …") will not have every chemical annotated but will likely have the term Pesticides in the annotations.

Often, too, annotations will have subheadings of interest. In the realm of chemicals, these subheadings can indicate whether the toxicity (adverse effects, poisoning) of the chemical is being discussed or the therapeutic use of the chemical.

The MeSH Mine sheet pulls out substances from the MeSH terms, finds their associated entities and writes them with counts and flags to the MeSHMine sheet.

Here we'll look at how the MeSH mining function works. To start, let's go to PubMed and find the article with PMID 28178360. Here's a link: [Melatonin protects oocyte quality from Bisphenol A-induced deterioration in the mouse - PubMed (nih.gov)]

Read the abstract and then on the right, click on MeSH Terms. You'll see this:



You'll see a variety of MeSH terms at the top of the list: substances, species, organs, processes, anatomy, cells, etc.

The substance list at the bottom is what the MeSH mine function focuses on. The Abstract Sifter extracts those terms to put on the MeSHMine sheet.

Figuring out whether the substance is discussed therapeutically or for its adverse effects is slightly more complicated. The Abstract Sifter code looks for subheadings or qualifiers associated with a substance (or its parent family term).