INTERNET SEARCHING: GOOGLE & GOOGLE SCHOLAR

Google (www.google.co.uk)

What is included in Google?

Google covers all web data crawled and retrieved by Google's Googlebot.

Google is currently the most-used search engine. However, no search engine searches the whole of the web. If it is important for you to find *all* results, or you get nothing on Google but think that there ought to be a result (or more results), then repeat your search on another search engine.

Alternatively, use one of the metasearch engines, which search multiple search engines simultaneously.

Metasearch websites:

- www.dogpile.com
- www.metacrawler.com

How are Results Ranked?

By default, Google search results are ranked with the results Google feels are most relevant first. The algorithm by which results are sorted is proprietary, so nobody outside Google knows exactly how it works. However, we do know that it uses a combination of factors, including but not limited to:

- Secured site (https, not http)
- Mobile-friendly
- Speed of loading
- Content quality
- Content length
- Quality backlinks (links *to* the page from other good websites)
- Domain age
- Traffic

So a website with a large amount of high-quality content on a website that has been around for a number of years, is accessed frequently and is linked to from other quality websites is likely to rank high.

However, it is possible to force results to be ranked by date of last update (see above).

How up to date is Google?

Googlebot crawls websites 'regularly' and updates its record of the content. Depending on the nature of the site, updates could occur at any time between days or weeks. The larger and more active a site is, the more likely that it will be updated quickly.

Searching Google

Word Choice

Google simply crawls fulltext; it does not work on assigned keywords. Therefore, think like the author: what words do you think they would have used in the sort of document you are looking for?

Google uses both stemming and synonyms.

'Stemming' is where the word stem is used for the search, so if you search *'teen'* you will get results for *teen, teenager, teenagers,* etc. This happens automatically.

Synonyms are more complex; Google's algorithm looks at your search term and will automatically include results with words it thinks are synonyms. So a search for 'dose' is likely to also include results for 'graphics' or 'photographs'. The algorithm is quite good at only picking relevant synonyms, but not 100% effective.

Therefore:

- If you want to search for an exact word, and 'turn off' stemming and synonyms, put the word in double quotation marks: *"picture".*
- If you want to be sure that a particular synonym is included, use OR: *picture OR photograph.*

Refining Your Search with Operators

Double quotation marks around a phrase	Searches for that precise phrase	"risk management of medicines stored in clinical areas" This also works to force Google to search for a single word exactly, rather than (as it usually does) returning results including words Google regards as similar.
Combine searches	Put OR between your search terms to search for pages with any of two (or more) terms. Without 'OR', Google will search only for pages with both terms. 'OR' must be in capitals.	- Pharmacist OR "pharmacy technician" will find pages with the word 'pharmacist' and the phrase 'pharmacy technician' - *Pharmacist "pharmacy technician"* will find pages with the word 'pharmacist' *and* the phrase 'pharmacy technician'.
Wildcards	Use an asterisk * to substitute for a *whole word* in a search phrase. You can use more than one asterisk in your search phrase. Note: you cannot use this to substitute for part of a word.	Searching for *what is the biggest hospital in ** will bring back search terms referring to the biggest hospitals in any city, country, etc.
Site:	Searches only within that particular website, or domain.	- *Temperature control site:nhs.uk* searches for mentions of web pages including the words temperature and control, but only on NHS sites - *temperature control site:sps.nhs.uk*

		searches for pages with the words temperature and control, but only on the www.sps.nhs.uk website.
-	Excludes words from search	*Tablet –computer* will search for pages that mention the word 'tablet' but not the word 'computer'
Intitle:	Searches only in the title	*intitle:"high dose loperamide"* will search for pages with the phrase "high dose loperamide" in the title.
Filetype:	Searches only for files of the specified type.	*filetype:xls "list of UKMi"* will search for Excel documents with "list of UKMi" in them.
Related:	Searches for sites/pages Google considers related to the one specified.	*related:sps.nhs.uk*
Inurl:	Searches only in the URL (web address)	*inurl:"clinical governance working group"* will search for the phrase "clinical governance working group" in the page URL.

Notes:

- Search operators can be combined to construct more complex searches. However, Google has a limit of 32 words for a search.
- Google does not recognise brackets in searches.

Natural Language vs Search Strings

Google now understands natural language questions. So you can do either of the following:

- *Maximum dose flucloxacillin*
- *What is the maximum dose of flucloxacillin?*

In most cases, however, searching by words you expect to find in your target document is the most appropriate way to do it as this allows you more control over how Google interprets your search.

Additional functions

By default, Google shows all its results. However, you can choose to see only:

- Results from news sites
- Images (can be useful for tablet identification)
- Videos
- Maps
- The 'more' button allows you to select 'Books' for results from Google Books.

Tools

On the results page, under the main search box, there is a link for 'tools'.

- If you use the www.google.co.uk rather than www.google.com, this allows you the option to limit your results to UK results only.

- 'All results' can be changed to 'verbatim'. Verbatim means that Google will *only* search for the terms you input; it removes personalised, corrected, suggested, related, and non-inclusive results.
- 'Any time' can be changed to filter results between dates you choose.
  - Google displays results in order of relevance by default. However, if you limit your search results to a timeframe, you then have the option to sort by relevance (the default) or sort by date.

Cached Pages

A 'cached' page is a temporary copy of a webpage. It can be stored either on the client computer (yours) or elsewhere (e.g. at Google). Cached pages are stored to reduce lag time.

- On your computer/in your browser, if you press the 'browser back' button, the browser may show the cached page rather than displaying the page 'new'. This is more efficient as it does not require the browser to go back to the website and display new information.
- Elsewhere, cached pages allow pages to be displayed quickly at times of high workload, or if the website is down.

If you want to make sure that you are *not* seeing a cached page, 'refresh' the page.

If you want to see a cached page (which you may, if you suspect that information has changed), then Google allows you to do so by clicking the down-arrow at the end of the green web address line on search results (below the blue title hyperlink). Clicking the down-arrow reveals another menu which allows you to select 'cached'. Clicking on this takes you to Google's last cached version of the page.

Similar Pages

For some web pages (not all) Google will show pages it thinks are similar, if you wish. To do this, click on the down-arrow at the end of the green web address line on search results, as for viewing a cached page. If Google thinks there are similar pages, 'similar' will also be an option you can choose.

Google Advanced Search

If you find it difficult to remember the operators, or you are trying to construct as complex a search as Google allows, you may wish to use the Google Advanced Search page, which gives you boxes to fill in depending on which operators you wish to use.

Available at: https://www.google.co.uk/advanced_search

Alternatively, on the ordinary Google homepage (www.google.co.uk), go to the bottom of the page. There is a grey bar with Privacy / Terms / Settings in the bottom-right corner. Click on Settings, and this will give you several options, one of which is Advanced Search.

Google Alerts

If you have a particular topic of interest, you can set up a Google Alert that will send you an email (daily or weekly) when new information is published. This can either be a long-term interest, or short

term – for example, if you are following the development and arrival on the market of a particular medication, or you are waiting for a document on a particular subject to be published.

Google Scholar (https://scholar.google.com/)

Google Scholar (GS) is a search engine intended for academic research; it can best be described as Google, but only including results Google thinks are 'scholarly'. The search functions are also slightly different to the ordinary Google interface. In a *Nature* survey in 2014, 60% of scientists said that they used GS regularly.(1)

What is included in Google Scholar?

In 2015, it was estimated that Google Scholar included 160-165 million records.(2) Unlike bibliographic databases like Medline or Embase, GS does not start with a list of journals and then index the articles. It starts by finding the articles on the web, and deciding whether the articles fit their criteria for 'scholarly'. So you will find research articles, conference abstracts, academic books, dissertations, and so on. It is less likely that you will find editorials. Articles without authors are often not included.

The definition of 'scholarly' is 'what scholars think is scholarly'; if an article cites, or is cited by, other articles GS thinks are scholarly, then GS will class the original article as scholarly. Although this sounds like a relatively dubious way of identifying scholarly content, it does work reasonably well in that more scholarly articles are linked more strongly into the 'scholarly web'.(1) It does allow the inclusion of 'grey literature' (see below), which can be either an advantage or a disadvantage.

Although Google Scholar covers a vast number of articles, coverage is not equal across all sciences and all languages. Coverage by Google Scholar varies between disciplines; 'hard' sciences tend to be better covered than social sciences and humanities. Coverage for biomedical sciences – including medicine and pharmacy – is generally good. However, some publishers have refused to allow Google to crawl their site, resulting in results from that site being absent from Google Scholar.(3)

Grey Literature

'Grey literature' is documents that are not published by commercial publishers (e.g. journal articles). It includes academic dissertations and theses, reports and papers produced by governments and other organisations, etc.(4) These documents are unlikely to be indexed in the major bibliographic databases, but may still contain useful information, particularly if the subject is either very new or is not the kind typically published in academic journals.

One group of researchers found that 8-39% of Google Scholar content was grey literature. However, GS does not include all grey literature. If it is important to find *all* such information, hand-searching of individual organisation websites is likely to be required.(4)

How are results ranked?

By default, Google Scholar ranks (sorts) articles by relevance. Its algorithm aims to evaluate 'relevance' by looking at the fulltext of each document, where it was published, who the author is, and also by how often and how recently the article has been cited in other scholarly literature.

You can change the sort order to sort by date (most recent first) by clicking 'sort by date' in the menu on the left-hand side.

How up to date is Google Scholar?

New papers are added to Google Scholar several times a week, whenever Google's search robots find new material. Since this relies on the robots crawling the web, it is likely that articles on bigger websites with more traffic (e.g. BMJ) will be added faster than those from websites with less traffic.

Updates to existing records take longer – up to a year or more – because the robot has to recrawl the source website. For some of the larger websites, the speed of update is limited by the rate at which the website allows the Google robots to crawl the site.

Searching Google Scholar

Unlike Embase and Medline (and other such bibliographic databases), Google Scholar does not assign keywords to articles. Like Google, it searches the fulltext of the article itself. This can be either an advantage or a disadvantage.

It can be an advantage if you know what words the author is likely to have used in the article, but do not know what keywords a bibliographic database would assign, or if there are limited number of highly-relevant words that are likely to appear in the articles you want to find.

On the other hand, if your articles are not likely to contain unique words, searching via Google Scholar is likely to be more difficult than searching via a bibliographic database with a limited number of assigned keywords.

In short, when searching Google Scholar, you have to think like the author; when searching a bibliographic database, you have to think like a keyword-assigner.

- Operators are used as for Google.
- Searches are limited to 256 characters
- Only the first 1000 results are shown.
- 'Nesting' terms in brackets, e.g. (pharmacist OR technician) AND (hospital OR community) are not recognised by Google Scholar.

Google Scholar Advanced Search

Google Scholar Advanced Search is available via a popup-box over the ordinary search page. To find it:

1. Click on the three parallel lines on the top-left corner of the page.
2. Click on 'Advanced Search'.

Google Scholar Advanced Search allows you to find articles:

- With all of a series of words in random order
- Containing an exact phrase
- With at least one of a list of words
- Without certain words

You can specify that your search terms appear anywhere in the article, or only in the title of the article.

- Articles authored by a particular name
- Articles published in a particular journal
- Articles dated between certain dates.

Some of this functionality does not require the Advanced Search screen; you can do it via the normal Google Scholar.

Citation Tracking

Identifying Citations

Google Scholar searches the whole of the web pages and documents it crawls. It can therefore identify citations which are mentioned on a web page (e.g. the reference list of an academic article) and include these in its search results list. Therefore, there is the possibility for a citation to appear in a Google Scholar search that does not correspond to any full document Google has located.

However, it should be noted that not all of what GS identifies as a 'citation' is what we might regard as a citation, as GS cannot differentiate between the citation of a journal article and a student handbook or library guide.(3)

Cited By

For each document Google Scholar retrieves, it will also link to all the documents citing that document. The list can be accessed by clicking the 'cited by' link beneath each search result. This is very useful if you have found an article that is relevant but old, as it allows you to track more recent research.

Highly-Cited Articles

Highly-cited articles – those cited by large numbers of other articles – are often the most important and/or influential work in the relevant area. Not only can Google Scholar identify these, but it routinely uses the number of citations to rank search results.(5)

Google Scholar 'My Library'

To use the Google Scholar 'My Library' effectively, you need to be signed in to Google. If you do not have a Google account, you can sign up for one (free) at:

https://accounts.google.com/signup/v2/webcreateaccount?hl=en&flowName=GlifWebSignIn&flowEntry=SignUp

'My Library' allows you to save links to articles of interest so that you can read them later.

- To save an article: click on the star underneath each search result. The star will turn blue.
- To access your saved articles: click on the 'My Library' link on the top right hand corner of the page.

In the My Library page, you can manage your results. Tick the box by the result you wish to manage, and:

- Edit the way the result shows in your library by clicking the pen icon at the top of the page.
- Remove the article from your list by clicking the dustbin icon.
- Export the citation to reference management software (BibTeX, EndNote, RefMan, or CSV) by clicking the down-arrow icon.
- Add a label by clicking the label icon. If you have no labels yet, you can also create a new label here, or manage existing labels.

Getting Citations

To get a citation for an article you are interested in, click the " (speech marks) button beneath the search result. This will bring up a pop-up window with several options – you can cut and paste the citation in one of the several styles it offers, or send the citation to your reference management software (BibTeX, Endnote, RefMan, RefWorks – if you use Zotero, then you don't need to use this step, just click the yellow folder icon in the top bar of the browser as usual, and select the articles you wish to save, as usual).

Labels

Labels are a method for dividing your My Library contents into subject categories. You can add as many labels as you like to a search result.

Once you have some labels, a list of them will appear down the left hand side of the page. At the bottom of the list, there is a link for 'Manage labels'; here, you can add new labels, edit, or delete labels.

This is useful if you wish to keep the results of several different searches in My Library, as you will be able to filter your My Library contents to see only those contents with the selected label.

Linking to your Trust Library

Google Scholar has the option to link to your local academic (including NHS Trust) library. This will give you a link directly to the fulltext via your library's usual route.

To do this:

1. Click on the three horizontal lines in the top left-hand corner of the page.
2. Click on 'settings'
3. Click on 'library links'
4. Use the search box to search for your library. You can choose up to five libraries, e.g. "fentanyl patch"
5. Select the libraries you want.
6. Click 'save'.

Note: because Google Scholar indexes papers automatically, sometimes it will get the metadata wrong. If this is the case, it may not be able to link to an article in your local library's holdings successfully. Therefore, if an article you are particularly interested in is not marked as held by your local library, it is still worth checking via Athens/your library website.

Google Scholar Button

The Google Scholar button is a web extension for your browser (including Chrome, Firefox, and Microsoft Edge – you can download it from the relevant web store). When you are on the page of a scholarly article, it displays the Google Scholar result for that article, including:

- A button to retrieve a citation (useful if you want to cite the article but are not using a reference manager, or don't want to go via your reference manager)
- A link that will take you to a Google Scholar list of all articles citing the original article
- A link for a list of related articles
- A link for a list of different versions/pages for the same article
- A button to take you to a free fulltext article, if available.

If you are on a page with multiple results, you can select (highlight with your cursor) one of the results, then press the Google Scholar button, and you will get the above information for that article.

Google Scholar Alerts

If you have a topic of particular interest, and wish to be notified when new research is published, you can set up a Google Scholar Alert. This works in a similar way to Google Alerts, except just for Google Scholar results.

Google vs Google Scholar vs Bibliographic Databases

What are they good for?

- **Google** is good for finding general information. If you know very little about a subject, then Google is a good place to start to get some basic background information and learn what vocabulary is being used to discuss the subject.
- **Google Scholar** is good for:
  - Wide searching, as its coverage is likely to be wider than any single bibliographic database.(6)
  - 'Quick and dirty' searches when you just want a general picture of what is being published in the scholarly literature, or you just want to see if anybody is writing about it at all.(7)
  - When you want to track citations.
  - When you are dealing with a subject that doesn't match up to bibliographic database keywords very well, and a text search of fulltext articles will be more likely to gain good results.
  - When you want to include 'grey' literature.

- o Looking for fulltext articles. GS may identify more fulltext articles than bibliographic databases.(7)
  - o Identifying highly-cited 'important' papers, as the number of times an article is cited is one of the factors GS uses to determine its position in search rankings. Furthermore, each search result has a 'cited by' link, with the number of times the article has been cited so it is possible to see which are the most-cited articles in a list.
  - o Finding related articles, as each article search result has a link for 'related articles'.
- **Bibliographic databases** are good for:
  - o When you need to do a complex search with limits that Google Scholar does not allow.
  - o When your search is likely to be developed incrementally, as Medline and Embase allow you to build searches line by line, but Google Scholar does not.
  - o When your search subject is one where it is an advantage to search keywords rather than fulltext, for example if your topic is likely to be referred to in passing in non-relevant articles, or your most useful search term is also used in another way (e.g. a surname).(8)
  - o When you want the comfort of a formal, structured search that will reassure you that you have found at least *most* of the important information.
  - o Precision. Assignment of keywords (when correct) makes it easier to construct searches with a high degree of precision.

What are they bad for?

- **Google** includes all the material crawled by Google. Much of it may be irrelevant, out of date, or unreliable.
- **Google Scholar**
  - o This relies on automatic 'crawling' of sites to pick up metadata. If the metadata is wrong, or the bot does not crawl the site, then the result will either be wrongly indexed or absent.
  - o Complex searches are difficult as Google Scholar does not have the range of limits that the bibliographic databases do.(6)
  - o Because keywords are not used, you have to search using words included in the articles you wish to find. This may be difficult if you are not familiar with the usual vocabulary used.
  - o Reproducible results. Google updates its algorithms hundreds of times a year, and regularly re-crawls websites. Therefore, the same search terms may not get the same results in the same order if repeated on a different day, or in a different location.(5)
- **Bibliographic databases**:
  - o These rely on manual addition of records to the database, so this may be slow.
  - o Keywords are added manually, so – given human error – they may be illogical, inconsistent, incomplete, or even absent.
  - o Bibliographic databases generally do not include 'grey literature'.

Note:

If it is essential that you find *all* of the information available on a subject, neither Medline, Embase, nor Google Scholar will be sufficient alone.(6,9) The best course of action is to search all three.

References

1. Van Noorden R. Google Scholar pioneer on search engine's future. Nature News [Internet]. [cited 2018 Aug 1]; Available from: http://www.nature.com/news/google-scholar-pioneer-on-search-engine-s-future-1.16269

2. Orduna-Malea E, Ayllón JM, Martín-Martín A, López-Cózar ED. Methods for estimating the size of Google Scholar. Scientometrics. 2015 Sep 1;104(3):931–49.

3. Al. JEGE. Scholarish: Google Scholar and its Value to the Sciences. Issues in Science and Technology Librarianship; 2012.

4. Haddaway NR, Collins AM, Coughlin D, Kirk S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLOS ONE. 2015 Sep 17;10(9):e0138237.

5. Martin-Martin A, Orduna-Malea E, Harzing A-W, Delgado López-Cózar E. Can we use Google Scholar to identify highly-cited documents? Journal of Informetrics. 2017 Feb 1;11:152–163.

6. Halevi G, Moed H, Bar-Ilan J. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature. Journal of Informetrics. 2017 Aug 1;11(3):823–34.

7. Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, et al. Retrieving Clinical Evidence: A Comparison of PubMed and Google Scholar for Quick Clinical Searches. Journal of Medical Internet Research. 2013 Aug 15;15(8):e164.

8. Shultz M. Comparing test searches in PubMed and Google Scholar. J Med Libr Assoc. 2007 Oct;95(4):442–5.

9. Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Systematic Reviews. 2016 Mar 1;5(1):39.

Author

Jen Smith MRPharmS, (West Midlands Medicines Information Service), on behalf of the UKMi Clinical Governance Working Group.