UKMi Quality and Risk Management Group

INTERNET SEARCHING FOR MEDICINES INFORMATION STAFF

Version 2.0
March 2020

Contents

Introduction

Although – given how much we use the internet in daily life – it is tempting to assume that we already know how to use it, most of us only learn the basics at school and even at university. In some cases, we may even have been taught that it is not appropriate to use general internet searching – e.g. via Google – for answering medicines information enquiries, and only bibliographic databases (e.g. Embase, Medline) should be used.

Although there are many pitfalls in using a general internet search as part of an enquiry answering strategy, there are also many advantages.

The purpose of this document is therefore to provide guidance on the best use of the internet and internet-based search tools for enquiry answering in Medicines Information. It is not intended to be a detailed manual on all aspects of internet use.

Summary

- A general internet search can be a very useful tool in answering MI enquiries; it has advantages as well as disadvantages compared to the more usual searches of bibliographic databases (e.g. Medline, Embase). As with every resource used in MI, it should be used with consideration for its strengths and weaknesses, and together with other resources as appropriate.

- General internet searching is a different skill to use of bibliographic databases.

- The biggest and best search engine is Google. This is therefore the first choice for use in MI.

- Alternative search engines (e.g. DuckDuckGo, which does not personalise results) may also be used, but Google is the first port of call.

- 'Private Browsing' does not prevent personalisation of search results.

- Some websites may show up better in certain browsers (Internet Explorer is particularly idiosyncratic).

- Google has a number of useful tools (e.g. Google Advanced Search, Google Scholar) with which MI staff should familiarise themselves.

The place of a general internet search in enquiry answering strategy

It may be tempting to assume that a general internet search should always come last, after all of the other 'conventional' MI resources have been exhausted. However, this is not always the case.

The advantage of the standard MI resources is that most of them cover a specific area and can be searched in a standardised way. You already know what sort of information will be found in each resource and you will only look in the resources likely to hold the information you want.

A general internet search is useful:

- When you would otherwise search a bibliographic database (e.g. Embase/Medline) but the most appropriate search terms aren't in their thesauri.
- If the keywords you use are not listed as keywords on Embase/Medline, and are not used in the title/abstract of the article listing (or anywhere else in the Embase/Medline record). Because a search engine searches the fulltext of an article (if available), you may get additional relevant results.
- When you want the latest information. As Embase/Medline require a human being to fully index records, there may be a time lag between the publication of a paper and its appearance in Embase/Medline. An internet search is likely to find new papers faster, especially from the more influential journals with larger, more active websites.
- When you just want to know "is this a thing?" rather than finding *all* the literature on a subject.
- When you are looking for guidelines, consensus statements, expert opinions, etc, that are not published in the academic literature or other 'grey' literature.
- Identification/availability of products, e.g. branded products, foreign products, herbal/complementary products.
- Information on drugs of abuse/substance misuse.

As with any other resource, the key to effective use is knowledge: knowing *how* to use it, and also *when,* making best use of the advantages and avoiding the pitfalls.

Internet: the Basics

Glossary

Bookmarks/Favourites: most browsers provide a facility for users to save links to frequently-visited sites to avoid having to type the address into the search bar every time.

Browser: software used to view web pages, e.g. Google Chrome, Microsoft Edge, Firefox.

Domain name: the part of a URL (web address) that is intended to give information on the type of website and its origin.

Homepage: the 'front page' of a website where the site's designer intends that people visiting the site should arrive first.

Internet: a global system of interconnected computer networks.(1)

Search engine: a software system designed to search the World Wide Web for information in response to a user query.

URL: "Uniform Resource Locator" – also known as a "web address" is the unique internet address of a web page, e.g. www.sps.nhs.uk

World Wide Web: a global collection of documents and other resources, linked by hyperlinks and URLs.(2)

Domain Names

The domain name is a website's basic address, which usually goes to its homepage, e.g. sps.nhs.uk. This is composed of various parts: the site-specific part, the top-level domain, and possibly also a country code.

Top Level Domain (TLD)

This is the domain extension, e.g. .com, or .org. It is intended to provide some information about what type of website this is:

- .com was originally intended for for-profit business sites, but is now used for all kinds of sites. (.co.uk is the UK equivalent).
- .org is for non-profit organisations (.org.uk is the UK equivalent).
- .gov is for government sites. (.gov.uk is the UK equivalent).

There are many more TLDs nowadays than in the early days of the web (e.g. .pub for bars and pubs, although some publishers use it too!).

The NHS has its own TLD, which only NHS organisations are permitted to use: .nhs.uk.

Country code TLD

These are suffixes which are intended to indicate the country in which the website is registered, or the nationality of the registrant (different countries have different rules for this). For example, .uk for United Kingdom sites.

Many USA sites do not use the .us suffix; the absence of a country code may indicate that the site is American (or may not).

There are more country codes than there are official countries: for instance, there are codes for the European Union (.eu) and Antarctica (.aq).

See: https://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Browsers: Differences and Choice

A *browser* is the software you use to view web pages: your window into the internet. There are several browsers available; they each have their advantages and disadvantages.

The most commonly used are:

- **Internet Explorer** (Microsoft) is used widely in the NHS, as it is the browser which comes bundled with Microsoft Windows software (until Windows 10). However, it can sometimes be idiosyncratic, and some websites will not work properly with it. If this is a problem, use an alternative browser, e.g. Chrome or Firefox.
- **Microsoft Edge** (Microsoft) is supplied with Microsoft Windows 10 as a replacement for Internet Explorer. As of January 2020, its inner workings will be based on Google's open source Chromium code. It is likely, therefore, that it will have fewer website-compatibility issues. It will be available for MacOS as well as Windows-based computers.
- **Chrome** (Google) is Google's browser, and is built on Google's open-source Chromium code. It works on any platform and is part of the Google ecosystem, so convenient to use for people who have a Google account. It is the most widely-used browser in the world and is very fast, but tends to use significant amounts of your computer's memory. Therefore, if you are using an older computer and experiencing slowdown issues, switching to a different browser (e.g. Firefox) may help.
- **Mozilla Firefox** (Mozilla Foundation). Firefox is built on Mozilla's open-source code, and works on any platform. The newest version is as fast as Chrome, but less memory-hungry.
- **Safari** is only for Apple devices, so unlikely to be encountered in Medicines Information Centres, although many people will use it on their personal devices. It is a capable browser, and has a significant focus on privacy and security. It is not as fast as Chrome, but is deeply integrated into Apple software systems, so battery life on devices tends to be good, and the integration can provide productivity advantages which outweigh the slightly slower experience.

Private Browsing

Most browsers offer some form of 'private browsing'. Many people believe that this protects them from being tracked online, but this is not the case.

Normally, as you move through the internet in an ordinary browser session, your browser will keep a history of where you have been (in case you want to go there again), cookies from websites, and data you input into forms so you can autocomplete later. All of these aspects make your online time a little easier and more convenient. However, there may be times when you do not want your browser to save this data, for security or privacy reasons – particularly if you are entering sensitive data onto a shared computer.

The answer to this is 'private browsing' (also known as 'Incognito Mode' on Chrome), where the browser will not save this data after you close the session.

What 'private browsing' does *not* do is hide your identity or location, or prevent the internet sites you visit (including search engines) from tracking you or collecting your personal data.

Cookies

HTTP cookies (web cookies, internet cookies, etc.) are not baked goods. A 'cookie' is a small piece of data which the website sends to your computer, and which your computer stores in your web browser. They were originally (and still are) used for essential functions such as remembering the items in your internet shopping basket, and knowing whether you are logged into a website or not. However, cookies can also be used to track internet usage and thus compile a long-term record of a user's internet use history.

As this raises privacy concerns EU law requires that websites which download cookies to the user's browser get 'informed' consent for this, usually via a pop-up box which the user must interact with in order to get it to disappear.

Search Engines: Differences and Choice

A *search engine* is software designed to search the world wide web for information in response to the user's query.

As the world wide web comprises billions of web pages, this creates several problems:

- How does the search engine know the page exists?
- How does the search engine know that the page is relevant to the search?
- How does the search engine sort the results of the search into a useful order?

Websites are added to a search engine's results automatically (rather than having to be manually submitted) when the search engine's 'spider' (software which is designed to find new websites and evaluate their place in the web) 'crawls' the web and finds it.

Once a search engine knows about a webpage, it will evaluate its content by finding keywords (used either openly in the webpage itself, or in the metadata 'behind' the web page.

- **Google** is the search engine with the largest coverage and the most advanced algorithms. Therefore, Google is the first-choice search engine for use in medicines information. No general internet search should be regarded as complete without using Google. However, Google tracks user activity (even when you are not logged in) in order to 'personalise' your search and present you with results which it thinks you want. Google may, of course, not always guess correctly. This is likely to be less important when searching for non-political, factual, medical information, than for more contentious content, but should be borne in mind.
- **DuckDuckGo** is a relatively small search engine focused on privacy. It does not collect any user data. Therefore, you can be confident that your results will be 'clean' and without 'personalisation'. However, as DuckDuckGo does not have the capability of Google, its place in the MI centre is that of second-line search engine for use if it is suspected that results obtained via Google are inaccurate or incomplete due to excessive use of personal/browser history data.

- **Bing** is Microsoft's search engine. It is less popular than Google, and although it has some differences, it does not appear to offer any unique advantage that would be useful to an MI service.
- **Yahoo** search is powered by Bing.

There are other search engines available. However, for Medicines Information purposes Google should be used first-line. DuckDuckGo may be used after a Google search if it is suspected that Google results are unreliable due to personalisation.

Metasearch Engines

Metasearch engines are like search engines of search engines. A metasearch engine searches multiple search engines at once; this may increase coverage, but as all search engines do not handle enquiries in the same way, it is best to keep search terms simple.

As the search engines searched by a metasearch engine may be small or obscure, a metasearch engine should be used *in addition to* a search of major search engine(s), rather than as a substitute for it.

Metasearch engines include www.dogpile.com

Saving Links and Web Addresses

There are a number of websites you will return to multiple times (e.g. those on the UKMi resource lists). It is inefficient to type the address into the browser search bar every time, so it is preferable to save the web address so that you only have to click on it.

The usual way to do this is to locate your browser's bookmarks/favourites functionality.

Once you get beyond a certain number of bookmarks, you will need to organise them into order; into folders is probably best, so you can keep all the 'adverse drug reaction' websites in one place.

It is useful to have a departmental list of bookmarks, which you can download and send to new members of staff when they join the department. Each browser will have a way to do this.

There are also bookmarking sites (e.g. Pinboard.in) which have additional, or different, functionality. You may have to pay for these.

Using the Internet

How to Search

Keyword choice

The internet does not have a standard thesaurus as the bibliographic databases do. Search engines look for words occurring in the text of websites in order to return results.

You should therefore attempt to think about:

- What words might the author have used in the article? Think about how you might have written an article on your topic of interest.
- What combinations of words might have been used?
- What words are *not* likely to be used?

This ability to search the fulltext of articles can be very useful if your query involves either a relatively unusual word, or a combination of unusual words. As search engines will search a whole document, you may get results that you do not get via bibliographic databases.

The disadvantage is if your keyword is one that can be used in two ways, e.g. tablet=pharmaceutical, or tablet=computer. In this situation, you are likely to have to use advanced searching techniques (see later) to eliminate irrelevant 'hits'.

Note that you may have to do several searches, using different keywords, in order to find the results you need.

See the section on using Google for more detail.

Advanced/Complex Searches

Some search engines (e.g. Google) have 'advanced search' capabilities which allow you to search for results that contain combinations of keywords, or do not contain certain words, etc.

Google's Advanced Search functionality is described later in this document; other search engines may not have such functionality or may operate differently. If you are not using Google, check how advanced searching works on the search engine you are using.

Natural Language Searching

Natural Language Searching is when you formulate your search query as if you were talking to another person: *What are the side effects of paracetamol* rather than as a list of keywords: *adverse effects paracetamol.*

Natural language searching requires a significant amount of computing power and excellent algorithms to achieve accurate results, as it requires the computer to understand how the human thinks, as opposed to keyword searching, which requires the human to understand how the computer thinks. As computers are simpler and more predictable than humans, the latter is much easier than the former. Advances have been made in natural language searching over recent years; even so, it is not advisable to use natural language searching in medicines information except as a last resort 'additional' search.

As natural language searching requires complex algorithms, it may work less well on the smaller search engines.

Evaluating Information Found on the Internet

The internet is not owned by anyone, and there is no restriction on what is put on the internet and by whom. As anyone can build a website with minimal technical knowledge, you will need to evaluate the information you find. The following criteria will help:

- **Evaluation:** Has it been evaluated by a trusted person/body already? This may be UKMi, or your organisation's library. Does the information appear to be accurate and trustworthy?
- **Reputation:** Is the resource well-known and well-respected?
- **Presentation:** Is the information professionally presented? Although professional presentation does not indicate trustworthiness, lack of professionalism may indicate that the resource is not to be trusted. Is the site well-designed and easy to use?
- **Author and Authority:** who wrote or commissioned the information? An individual, national association, professional body, commercial entity? What is their authority? Can you tell? Does the author have any relevant qualifications? Has the information been checked/reviewed by someone other than the author?
- **Audience:** Who is the target audience? Is it aimed at doctors (e.g. for use at point of care), pharmacists, nurses, or patients?
- **Referencing:** Is it referenced? Can you tell where the information has come from?
- **Date:** Is the information up to date? How frequently is it updated? Is the frequency appropriate for the type of information?
- **Scope:** What is the intended scope of the website? How broad is its subject reach, and how deep? Is it general, or specific to one narrow topic? Is the information fairly superficial, or is there more depth? Is the depth sufficient for your requirements?
- **Geography:** Is the information aimed at users in a particular country? For instance, medical information may include country-specific guidelines or reference to drugs only available in particular countries.
- **Bias:** Does the information appear to be objective, or is it presented with a particular bias?
- **Money:** Is there a charge for access to any part of the website? Is the website attempting to sell anything? If there is a charge, who receives the money?

QRMG has produced a **Resource Assessment Tool** for evaluating information sources, which is available at: https://www.sps.nhs.uk/articles/ukmi-recommended-resource-lists-and-tools/

Internet Forums, Discussion Groups and News Groups

Forums, discussion sites and news groups allow users – which may include anyone who wishes to contribute, or may require users to have certain characteristics (e.g. be working in an MI service) – to contribute information. This may be limited to asking questions (where questions are answered by employees of the site or other approved persons), or may also include providing the answers.

Some of these sites are aimed at a professional audience, and some are not. There is often little or no way to know how trustworthy the information on these sites is. Nevertheless, they can be useful as they can contain information not found elsewhere. For example, the lactation specialist Dr Thomas Hale runs a forum where people can ask about the use of medicines in lactation; he or one

of his staff will answer. As the site is moderated by his staff, the information on it is likely to be reliable. On the other hand, there may be no-one fact-checking information on patient forums. But even these can be a source of valuable information – e.g. reported side effects, or how patients deal with particular problems in practice. As with any other source of information, the information can be useful provided you use it appropriately.

Google Products in More Detail (www.google.co.uk)

Which Google?

Google is a single search engine, but it has multiple portals. Different portals offer advantages in different situations. For instance:

Country-specific Google

www.google.com is the standard Google homepage. However, there are other, country-specific, homepages which prioritise sites from those countries and allow you to easily limit results to those from the selected country, e.g.: www.google.co.uk

Google Scholar

Google Scholar returns only results which the algorithms regard as 'academic'. See later for more information, as this is very useful in MI.

Google Books

This is a search engine for the full text of printed books; how much content is displayed depends on arrangements with the copyright holder: for some books, only a small number of pages are available but for others, it is the whole book.

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 a metasearch engine, which searches multiple search engines simultaneously.

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.

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 'picture' 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.
OR	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 (i.e., Google assumes you mean to connect your search terms with AND unless you specify OR). 'OR' must be in capitals.	• Pharmacist OR "pharmacy technician" will find pages with the word 'pharmacist' OR 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:"quality and risk management group"* will search for the phrase "clinical governance working group" in the page URL.

Combining Searches

You can also combine searches using the operators above. For instance:

- *"pharmacy technician" OR pharmacist site:sps.nhs.uk* will search for pages that include the phrase "pharmacy technician" or the word "pharmacist" but only on the www.sps.nhs.uk website.
- *"pharmacy technician" OR pharmacist site:nhs.uk filetype:pdf* will search for PDF documents containing the phrase "pharmacy technician" or the word "pharmacist" hosted on any site with the .nhs.uk domain name.

Notes:

- Although you can combine search operators to produce complex searches, 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.

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. This gives you boxes to fill in for all the operators listed in the table above, depending on which you wish to use, plus the option to search for number ranges and other ways to limit the results.

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, manual 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.
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. It's very useful (with the site: operator – see above for information on operators) when you know or suspect that a specific website has information that is relevant to you but the site's own search function isn't able to find it for you.
- **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.
  - Looking for fulltext articles. GS may identify more fulltext articles than bibliographic databases.(7)
  - 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.
  - Finding related articles, as each article search result has a link for 'related articles'.
- **Bibliographic databases** are good for:
  - When you need to do a complex search with limits that Google Scholar does not allow.
  - 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.
  - 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)
  - 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.
  - 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. Be aware that documents located by Google may not be the most recent version, especially if they are hosted on a site that was not their original source (e.g. out of date UKMi documents re-posted by other NHS organisations).

- **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) (Also applies to the wider Google.)

- **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 Quality and Risk Management Group.