THEMATIC SERIES

How to write a scientific paper - Searching and Managing Biomedical information

Como escrever um artigo científico - Pesquisa e gestão da informação biomédica

T. Jacinto[a,*], **A. Morais**[b], **J.A. Fonseca**[a]

[a] *Biostatistics and Medical Informatics Department & CINTESIS, Faculdade de Medicina da Universidade do Porto, Porto, Portugal*
[b] *Pulmonology Department, Faculdade de Medicina da Universidade do Porto, Porto, Portugal*

Introduction

There is a massive and complex body of biomedical information, that keeps on growing daily. There are over 20 million references and nearly 5000 journals in Pubmed alone. For instance, the number of asthma related publications in MEDLINE rose from 586 in 1965 to an overwhelming 110351 in 2010. Moreover, health and life sciences represent most of the records and journals in ISI Web of Knowledge, with more items than exact, social and biological sciences.[1]

Nowadays, health professionals need to be able to recognize the need for new and updated information and knowledge, to identify and locate relevant sources, know how to access them, how to appraise the information and how to organize it. Biomedical literature tells us what is already known and helps prepare future work.[2] Furthermore, it is an essential component of the clinical decision, along with clinical experience and the values and preferences of patients.[3] There is evidence that immediately available online information can change and improve clinical decisions. Searching MEDLINE when inpatients are admitted significantly reduces costs and length of stay.[4] In a study by Crowley et al, 77% of 625 clinical questions of Internal

Medicine residents were answered using MEDLINE, which led to 47% of changes in clinical management.[5]

The aim of this paper is to address some topics on searching and managing biomedical information. We will discuss the construction of a search query and the development of a search strategy. We will use PubMed as an example, and demonstrate some of the tools that are provided. Hopefully, you will find some of the topics discussed useful when trying to find a specific answer to a question arising in your practice or engaging on a systematic search of data and information to access the current knowledge or ignorance in your medical field.[6]

Searching Biomedical Information

What to search?

You are searching for scientific publications, usually called the ''literature'', which are reports of empirical or theoretical scientific work. They can be very varied, such as journals, conference proceedings, books, technical reports, leaflets or patents. You will most frequently tend to be searching for scientific articles in journals, which are indexed in different online databases (based on various criterions, such as the existence of peer review, the scientific field and the periodicity of publications, amongst others). The first medical journal was the New England Journal of Medicine, with

---

* Corresponding author.
  E-mail address: tiagojacinto@med.up.pt (T. Jacinto).

Table 1   Top ten journals of General and Internal Medicine (Journal Citation Report, 2009).

	Impact Factor	Total Cites	Articles
New England Journal Of Medicine	47.05	216752	352
The Lancet	30.758	152843	280
Journal Of The American Medical Association	28.899	117090	234
Annals Of Internal Medicine	16.225	45184	174
British Medical Journal	13.66	71175	345
PLOS Medicine	13.05	8425	94
Annual Review Of Medicine	9.94	4257	36
Archives Of Internal Medicine	9.813	35977	203
Canadian Medical Association Journal	7.271	10024	98
Journal Of Internal Medicine	5.942	7470	105

its first article published in 1812.[7] Nowadays, MEDLINE alone indexes more than 5500 journals.[8]

It is possible to understand the influence of biomedical journals with the help of citation analysis. There are many techniques and indexes, but perhaps the most famous is the Impact Factor. The Impact Factor is the number of citations in the current year of items published in the previous 2 years divided by the number of substantive articles and reviews published in the same 2 years.[9] For example, if a journal has an Impact Factor of 5 in 2008, it means that the papers published in 2006 and 2007 had an average of 5 citations in 2008. The analysis of the Impact Factor should be made with caution, since there are several factors that can influence this index, such as the citation distribution of journals, the publication lag and the online availability of publications. These issues are further discussed by Dong et al.[10] Furthermore, you should not compare Impact Factors of different medical fields, because of different citation properties, such as citation density and half-life.[9]

You can find the annual reports of the Impact Factor and other indexes in the Journal Citation Report, provided by Thomson Reuters (more info at http://go.thomsonreuters.com/jcr). The top ten journals of the General and Internal Medicine, as of 2009, are shown in Table 1.

Search strategy

Before you begin your search, you should define a strategy. A search strategy is a set of simple actions that maximize the effectiveness of your search.

The first step is to frame a focused and consistent question to ask. A technique that aids the construction of the question is to use a structured framework. There are many frameworks, but the most referred is PICO.[11] PICO stands for Patient/Population/Problem, Intervention/Exposure, Comparison and Outcome, and can be extended to include the Type of question being asked (therapy, diagnosis, prognosis, etc.) and the Type of study design.[12] For example, you may ask if a shot of a steroid could work as well as five days of oral steroids in a young child after an asthma exacerbation. The question, based on the PICO framework, would be ''Among young children with acute asthma exacerbation (Population), is a single dose of IM dexamethasone (Intervention) comparable to five days of oral prednisolone (Comparison) for resolution of asthma symptoms (Outcome)?''.[11] In this case, you could also consider searching for Therapy studies, usually Randomized Controlled Trials.

Next, using the components outlined with PICO, define the concepts you want to include in your query, followed by synonyms and alternative terms. Delineate the specifics of your search: temporal interval (last 5 years, last decade, etc.), type of study (systematic reviews, randomized controlled trials, cohort studies, etc.) and participants characteristics (age, gender, etc.).

The best search strategies for a particular question are often achieved through trial and error, so keep on practising. The efficacy of your query will improve with time and experience.

Where to search?

There are many databases and search engines you can use. There are databases of specific types of studies, or specific diseases or conditions. You can find some examples in Table 2, adapted from Pai et al.[13] and Hull et al.[14]

PubMed has gained a high profile and authority in the biomedical field throughout the years. Because it is easy and quick to use and providing free access,[15] we will use it as an example. We believe that satisfactory search techniques are easily replicated in all search engines and databases. However, be warned that if you are conducting a systematic review, where search methods are fundamental to the validity of the study, using PubMed alone may not be sufficient, and you should include other databases and search engines.[13]

Pubmed was created in 1997 by the National Library of Medicine of USA, with the aim of providing free access to MEDLINE, a large biomedical database. It has over 20 million citations, 6000 journals, and 57 languages and is known for covering a large spectrum of clinical, biomedicine, bioethics and life sciences journals, covering publications since 1950.[16] You will not find the full text of the paper you are searching in PubMed, but instead, you will have external links to the full-text and also the possibility of navigating through related articles.

Searching Pubmed

PubMed has been recently updated, and now has a cleaner and more user-friendly interface. The search-box

Table 2 Examples of online databases and search engines.

Database/Search Engine	Domain	Size (# references)	Access
Pubmed (Medline)	Life sciences and Biomedicine	> 20.000.000	Free http://www.pubmed.gov
SciVerse SCOPUS	Broad scientific coverage	> 33.000.000	Requires Subscription (Elsevier) http://www.info.sciverse.com/scopus
ISI Web of Knowledge	Broad scientific coverage	> 15.000.000	Requires Subscription (Thomson Reuters) http://apps.isiknowledge.com
Google Scholar	Broad coverage	Not published	Free http://scholar.google.com
Embase	Biomedicine, Pharmacology	> 24.000.000	Requires subscription (Elsevier) http://www.embase.com
Cochrane Database of Systematic Reviews	Systematic reviews in health care	> 6600	Requires subscription (Wiley Online) http://www.thecochranelibrary.com
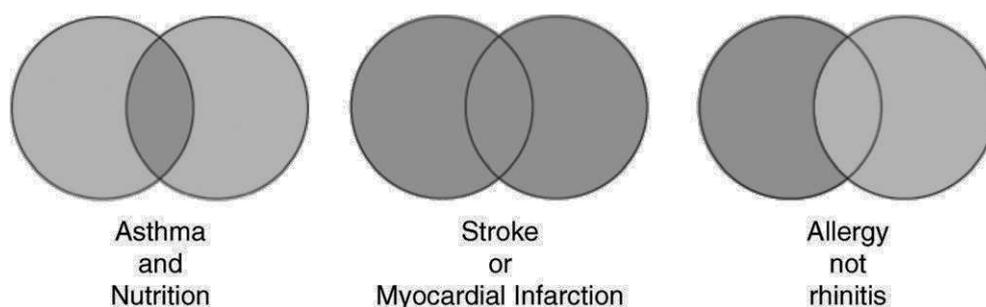
Figure 1   Boolean Operators
Dark grey area represent the output of your search, which is the result of the combination of the terms and the operators.

is available everywhere in the site, and you can use it whenever you want. Just fill in your query in the search box and click ''Go''.

You will most probably have several terms in your query, and you may combine them using special words. They are the Boolean operators, created by George Boole in the XIX century. You can use **AND**, **OR** and **NOT** (Fig. 1).

Using **AND** will narrow the results of your search, selecting only articles that contain all the terms linked with the operator, whereas **OR** will ''explode'' the search, increasing the number of results.[13] You can see some examples in Table 3.

The terms of your query can also be truncated, using a wildcard character, ''*''. For instance, if you search for ''Alerg*'', your output will include references with the terms ''Allergy'', ''Allergen'' and ''Allergic''. Moreover, there is a list of common words in the English language that are ignored when searching PubMed. They are known as ''Stopwords'', and the output generated when searching with these words would probably include all references available! They should not be included in your query.

Another useful feature of PubMed is that it is possible to specify a field in which a term should be applied. All references in PubMed are defined with several descriptors, such as Author, Journal, Publication Date, Title, etc. You can mark a term in your query with square brackets, so that it only applies to the specific descriptor. Table 4 shows examples of some descriptors and how to use them.

And of course, you can combine all these possibilities:

Wyatt J [au] AND BMJ [ta] AND last 5 years [dp] — Outputs all references with an author called Wyatt J, that have been published in the British Medical Journal in the last 5 years.

Asthma AND 2007 [dp] AND treatment [tiab] — Outputs all references with the term ''Asthma'', published in 2007 and which contain the term ''treatment'' in both the title and the abstract.

Table 3 Use of Boolean operators in search queries.

Query	Output
Asthma AND Nutrition	All references with the term ''Asthma'' and the term ''Nutrition''.
Stroke OR Myocardial Infarction	All references with the term ''Stroke'' or with the term ''Myocardial Infarction''
Allergy NOT rhinitis	All references with the term ''Allergy'' but not containing the term ''Rhinitis''
Myocardial Infarction AND (smoking OR obesity) AND mortality NOT diabetes	All references with the term ''Myocardial Infarction'' and the terms ''Smoking'' or ''Obesity'' and ''Mortality'' but not containing the term ''Diabetes''

Table 4 Examples of Search Descriptors.

Search Descriptor	Example
Author [au]	Wyatt J [au]
Journal title [ta]	BMJ [ta]
Language [la]	Portuguese [la]
Mesh Terms [mh]	Asthma [mh]
Publication Date [dp]	2007 [dp], last 5 years [dp]
Publication type [pt]	Review [pt], clinical trial [pt]
Title [ti]	Spirometry [ti]
Title abstract [tiab]	Environmental exposure [tiab]

PubMed also offers some useful tools, of which we describe the most relevant ones.

MeSH Terms

You can search PubMed using only keywords, but this method has some limitations. Using thesaurus-based searching may be more appropriate. Indexers assign descriptors to each reference, from a controlled list of subject terms. The terms used in PubMed are called MeSH Terms, acronym for Medical Subject Headings. There are over 23000 terms, updated weekly and reviewed annually. They have a hierarchical division, with narrower terms under broader terms.[17] For example, the term for Chronic Obstructive Pulmonary Disease is ''Pulmonary Disease, Chronic Obstructive'', and is organized under Lung Diseases, Obstructive > Lung Diseases > Respiratory Tract Diseases > Diseases. You can learn more about MeSH terms in a specific section of the website of the National Library of Medicine, available at http://www.nlm.nih.gov/mesh.

There are advantages when you use MeSH terms in your query. They link synonyms, grammatical and spelling alternatives together and represent unique meanings for homonyms words. Moreover, they should ensure more relevant references in the output of your search. MeSH terms facilitate the design of your query by allowing broader (''exploding'') or narrower searches (using ''subheadings''). The disadvantages are that the indexer attribution of a term may be wrong, leading to false positive results and also that recent concepts may not yet have a MeSH term.[2] There is no ''right'' way to use MeSH Terms or keywords, so try different combinations to see which one works best with your question.

Limits

With this tool, you can apply limits to your search, as you may have defined in your search strategy. You can state, amongst other things, the range of the date of publication, types of articles to be included, species (humans or animals), languages, study participant's gender and age group.

Clinical queries

Clinical queries are a set of instruments that add detailed components to your query, with the aim of improving your results. You can define if you only want to search systematic reviews or genetic studies. You can also include particular study types, such as etiology, diagnostic, therapeutics (default choice) or prognostic studies. Furthermore, you can say if you want your search to be more narrow and specific (fewer results but more accurate) or to be more broad and sensitive (more results but less accurate).

Journal Database

You can use this database to search for a particular journal. It is very useful if you want, for instance, to only retrieve articles from a given journal or analyze the publications of a journal in a specific medical field.

Single Citation Matcher

If you want to search for an individual article, it may be easier to use the single citation matcher. Just fill in the details of the reference you are interested in (journal, date, author name and title words). The more details you enter, the more accurate the search.

What to do with too many / too few results?

If you have too many results, it is probably best to refine the search terms. Use fewer synonyms or try to search only in the title of the references. Also, try to use Limits to restrict your search.

When you have too few or no results, search for errors in your query. Use broader terms and increase the timespan of your search. Try to use more concepts and terms, while removing terms that are very specific.

Getting the full-text

When you open a reference in PubMed, we will most probably find links to external sites where you can download the full-text of the article. They can refer to libraries, the website of the journal, or to online open-access repositories, such as PubMed Central (www.ncbi.nlm.nih.gov/pmc). Some journals or repositories will require a subscription or a single ''pay-per-view'' to give you access to the full-text of the article. However, most academic and research institutions have access to relevant databases and journals, and you can try getting the full-text through them.

Managing results

There are many methods that you can use to store the results of your search. You can write them in your notebook (not advisable!), save them in a text file or a spreadsheet, or use specific software called ''Bibliographic Managers''.

Bibliographic Managers offer many features to save and use bibliographical references, and are usually very flexible and versatile. They can be used to store references to items in many different formats and material types, search and output references in many citation styles, import references directly from online databases, insert references in a document and generate bibliographies, and also store links to documents related to the references.[18,19]

Bibliographic managers are available in many formats and have a long list of features, so you should try them (most are free and easy to install) to see which applies best to your workflow.[20] There are commercial, free, open-source and web-based options. Probably the best known commercial software is Endnote (http://www.endnote.com) or Reference Manager (http://www.refman.com). Other options include Mendeley (http://www.mendeley.com) and Zotero (http://www.zotero.org), both free. For web-based bibliographic managers, you can try CiteULike (http://citeulike.org) or Connotea (http://www.connotea.org).

Conclusion

Being able to search and identify relevant data and information is essential for the clinical decision and for clinical research. There are a number of rules and concepts, that are easy to apprehend and to use, and which considerably improve the outputs of your search.

As Isaac Newton said, ''If I have seen further it is by standing on the shoulders of giants''. In the Internet era, there are many ''giants'' of information available for you to use. Carefully define your aims and design your question, and with some practice and experience, you will find your ''needle in a haystack''.

References

1. Jacso P. As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. Current science. 2005;89:1537—47.
2. Roberts P, Priest H. *Healthcare Research: A Textbook for Students and*. Practitioners: Wiley; 2010.
3. Tonelli MR. Integrating evidence into clinical practice: an alternative to evidence based approaches. Journal of evaluation in clinical practice. 2006;12:248—56.
4. Klein MS, Ross FV, Adams DL, Gilbert CM. Effect of online literature searching on length of stay and patient care costs. Academic Medicine. 1994;69:489.
5. Crowley SD, Owens TA, Schardt CM, Wardell SI, Peterson J, Garrison S, et al. A Web-based compendium of clinical questions and medical evidence to educate internal medicine residents. Academic Medicine. 2003;78:270.
6. Greenhalgh T. *How to read a paper: The basics of evidence-based medicine*. BMJ books; 2010.
7. Warren J. Remarks on angina pectoris. The New England Journal of Medicine, Surgery and Collateral Branches of Science. 1812;1:1—11.
8. MEDLINE Fact Sheet. National Library of Medicine; 2011 [24/04/2011]; Available from: http://www.nlm.nih.gov/pubs/factsheets/medline.html.
9. Garfield E. The history and meaning of the journal impact factor. JAMA: the journal of the American Medical Association. 2006;295:90.
10. Dong P, Loh M, Mondry A. The ''impact factor'' revisited. Biomedical digital libraries. 2005;2:7.
11. Moyer V. Weighing the Evidence: PICO Questions: What Are They, and Why Bother? AAP Grand Rounds. 2008 January 1; 19:2.
12. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC medical informatics and decision making. 2007;7:16.
13. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, et al. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. The National medical journal of India. 2004;17:86—95.
14. Hull D, Pettifer SR, Kell DB. Defrosting the digital library: bibliographic tools for the next generation web. PLoS computational biology. 2008;4:e1000204—10.
15. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB journal: official publication of the Federation of American Societies for Experimental Biology. 2008;22:338—42.
16. PubMed®: MEDLINE® Retrieval on the World Wide Web. National Library of Medicine; 2010; Available from: http://www.nlm.nih.gov/pubs/factsheets/pubmed.html.
17. Medical Subject Headings (MeSH®). National Library of Medicine; 2011; Available from: http://www.nlm.nih.gov/pubs/factsheets/mesh.html.
18. Mead TL, Berryman DR. Reference and PDF-Manager Software: Complexities, Support and Workflow. Medical Reference Services Quarterly. 2010;29:388—93.
19. Fourie I. Personal information and reference management - librarians increasing creativity. Library Hi Tech. 2011;29:0737—8831.
20. Fenner M. Reference Management meets Web 2.0. Cell Ther Transplant. 2010;2, doi:10.3205/ctt-2010-en-000087.01, e.000087.01.