Improving Efficacy of PubMed Clinical Queries for Retrieving Scientifically Strong Studies on Treatment

Salvatore Corrao, MD, Daniela Colomba, MD, Sabrina Arnone, MD, Christiano Argano, MD, PhD, Tiziana Di Chiara, MD, Rosario Scaglione, MD, Giuseppe Licata, MD

Abstract The authors evaluated the retrieval power of PubMed “Clinical Queries,” narrow search string, about therapy in comparison with a modified search string to avoid possible retrieval bias. PubMed search strategy was compared to a slightly modified string that included the Britannic English term “randomised.” The authors tested the two strings joined onto each of four terms concerning topics of broad interest: hypertension, hepatitis, diabetes, and heart failure. In particular, precision was computed for not-indexed citations. The added word “randomised” improved total citation retrieval in any case. Total retrieval gain for not-indexed citations ranged from 11.1% to 21.4%. A significant number of Randomized Controlled Trial(s) (RCTs) (9.1-18.2%) was retrieved for each of the selected topics. They were often recently published RCTs. The authors think that correction of the Clinical Queries filter (when they focus on therapy and use narrow searches) is necessary to avoid biased search results with loss of relevant and up-to-date scientifically sound information.


Introduction

PubMed is a web free resource and a powerful bibliographic citation search engine widely used by health-care professionals. It was developed and maintained by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) located at the National Institutes of Health. It covers many fields of biomedical knowledge: medicine, nursing, dentistry, veterinary medicine, the health care system, and preclinical sciences. At present, you can search more than 16 million bibliographic citations and abstracts, accessing both MEDLINE database and articles in selected life sciences journals not included in MEDLINE. All the articles in PubMed Central (free-full text journals) are included. PubMed offers clinicians the “Clinical Queries” page, a specific web page for clinical searches. You can access by clicking “Clinical Queries” from the PubMed sidebar. Then, you can choose one of the clinical study categories and enter your search term in the search box, selecting a scope. You can get a narrow and specific search or a broad and sensitive one. Clinical Study Categories tally with search filters based on the paper of Haynes and colleagues.

Clinical Queries are widely used by clinicians. If you focus your search on therapy and select a narrow search, you are joining your term onto a search string from the PubMed filter table as showed below:

(randomized controlled trial[Publication Type] OR (randomized-[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract]))

This is because randomized controlled trials (RCTs) are the sound study design to recognize strong evidences about therapy. However, the term “randomized,” used after the Boolean operator OR, may introduce a bias because it may cut off studies about therapy in which “randomized” could be exclusively written “randomised.” Haynes and colleagues, in the second part of the search string (after the Boolean operator OR), raise the need to find citations not indexed in PubMed. However, just in this case, not-indexed bibliographic citations could be lost when Britannic English had
Table 1: Comparisons between Search Strings Used to Retrieve Citations in Process, as Supplied by Publisher or Not-Indexed-for-MEDLINE, Each of Them Joined onto the Four Main Topics (PubMed Clinical Queries Vs. Modified Search String—see at the Bottom for Search String Syntax).

<table>
<thead>
<tr>
<th>Topic</th>
<th>String A (RCTA)</th>
<th>Number retrieved RCTs (RCTA)</th>
<th>Precision</th>
<th>Total retrieved citations</th>
<th>String B (RCTB)</th>
<th>Number retrieved RCTs (RCTB)</th>
<th>Precision</th>
<th>Total retrieved citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>62</td>
<td>44</td>
<td>0.71</td>
<td>74</td>
<td>52</td>
<td>0.70</td>
<td>18.2</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>95</td>
<td>71</td>
<td>0.75</td>
<td>109</td>
<td>80</td>
<td>0.73</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>Heart failure</td>
<td>45</td>
<td>32</td>
<td>0.71</td>
<td>51</td>
<td>36</td>
<td>0.71</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Hepatitis</td>
<td>27</td>
<td>22</td>
<td>0.81</td>
<td>30</td>
<td>24</td>
<td>0.80</td>
<td>9.1</td>
<td></td>
</tr>
</tbody>
</table>

RCTs = Randomized Controlled Trials
Var % = row variation percentage calculated according to the following formula: \[\frac{((RCT_B / RCT_A) \times 100) - 100}{RCT_A}\]

A: (randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract])) AND ("Premedline"[sb] OR publisher[sb] OR pubmednotmedline[sb]);
B: (randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] OR randomized[Title/Abstract]) AND controlled[Title/Abstract] AND trial[Title/Abstract]) AND ("Premedline"[sb] OR publisher[sb] OR pubmednotmedline[sb])

been exclusively used by authors. Moreover, we think that a bug in a search filter, though a little one, could twist clinician awareness of the searched topic. In our opinion, a slightly modified search string that includes the word “randomised” could avoid relevant information loss.

The aim of this paper was to evaluate the retrieval power of “Clinical Queries” specific/narrow search string about therapy in comparison with a modified search string that includes the word “randomised.”

Methods

The search strategy utilized by PubMed “Clinical Queries” is:

(randomized controlled trial[Publication Type] OR randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract]).

We compared retrieval efficacy of the string, mentioned before, to the subsequent one:

(randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] OR randomized[Title/Abstract]) AND controlled[Title/Abstract] AND trial[Title/Abstract]).

The proposed search strategy has the term “randomised” joined onto “randomized” by an OR (a Boolean operator) within round brackets. This term should improve recall of the PubMed string resolving a significant matter for clinicians: to reach right recall by the string used to narrow the search.

To compare the effects on recall of our change we tested the two strings joined onto each of four terms concerning topics of broad interest: hypertension, hepatitis, diabetes, and heart failure. Moreover, we used such modified search strategy to retrieve citations indexed for MEDLINE and, on the other hand, citations as supplied by publisher, in process, or not-indexed-for-MEDLINE. The search was performed on 28 December 2005 at 1:18 p.m. We calculated absolute and percentage variations of citations retrieved by the modified searching strategies compared with the PubMed ones. Moreover, we explored differences in total retrieved citations by both MEDLINE subset (citations included only in MEDLINE) and premedline, “as supplied by publisher,” and “not in medline” subsets. We used PubMed capabilities to modify search strategies to perform this kind of analysis using subset keywords. In the first case, we joined “AND medline [sb]” onto each string. In the second case, we joined on “AND (“Premedline”[sb] OR publisher[sb] OR pubmednotmedline[sb]).” Two independent observers (DC and SA) evaluated retrieved citations, and formally checked for randomised(ed controlled trials. A third person (SC) reviewed the entire process. Absolute and percentage variations of citations retrieved by the modified strings have been compared with the PubMed ones. Precision was computed, as reported by Haynes and colleagues, only for strings used to retrieve citations in process, as supplied by publisher, or not-indexed-for-medline. Moreover, we considered variations of the number of retrieved RCTs.

Results

The introduction of the word “randomised” improved total citation retrieval in any case. However, the main gain was achieved searching not-indexed-for-MEDLINE or in process citations. In this case, total retrieval gain ranged from 11.1% to 21.4%. Hypertension had the greatest gain among considered topics (19.4%). Searching total or indexed for MEDLINE citations, the gain was of about 1% compared with classic PubMed “Clinical Queries” strings. Table 1 shows comparisons between search strings used to retrieve in process, as supplied by publisher or not-indexed-for-MEDLINE citations, each of them joined onto the four main topics. Our data show that precision did not significantly change, but a significant number of RCTs was retrieved by modified string for each of selected topics. RCTs retrieval gain ranged from 9.1% to 18.2% (Table 1). Hypertension topic had the greatest gain also in this case. Often, Extra retrieved RTCs were recently published citations.

Discussion

Health care professionals widely use PubMed, one of the most relevant and powerful bibliographic citation search engines. Clinical Queries represents a clinician-oriented section of PubMed web site. Its scope is to make easier clinician task to search scientifically sound literature. However, some inaccuracies could act as confounders for unskilled users or bring a loss of information. Indeed, we found a bias in the Clinical Queries search filter that focuses on therapy and narrows the search. It does not merely consider the Britannic English variant of the term “random-
ized” (that is, “randomised”). We suggest a slight change in the search string by adding the term “randomised.” Our results point out a retrieval capacity of modified search strategy wider than PubMed Clinical Queries filter for retrieving scientifically strong studies on treatment. The modified search strategy was always able to retrieve a wider amount of citations and, in particular, of RCTs. These are methodologically sound studies that evaluate evidences about treatment. Notwithstanding, precision did not significantly differ between the two strategies (Clinical Queries versus modified string) joined onto each topic. When citations are in process, as supplied by publisher, or not-indexed-for-MEDLINE, PubMed filter has shown a significant loss of information. We think one could expect this observation. Indeed, the term “randomised” is utilized by journals using Britannic English like standard language. In this case, the PubMed filter systematically fails citation retrieval. On the contrary, “randomized” is always recognized by PubMed filter after NLM indexing process is completed (you have to consider that some citations will never be indexed). Then, up-to-date and relevant information can be lost focusing PubMed searches on studies about treatment by “Clinical Queries” filter. In conclusion, we think that clinicians have to be warned of the actual risk of losing relevant information when they focus PubMed search on therapy and select narrow search. Finally, in this case, correction of Clinical Queries filter is necessary to avoid biased search results with loss of relevant and up-to-date scientifically sound information.

References