Retrieving health literature is a cornerstone of evidence-based practice. With the rapid increase in available evidence, physicians can no longer rely on one or two key journals to stay current. Increasingly, physicians search bibliographic databases, such as PubMed, for research evidence, which is dispersed across hundreds of journals. Each year, physicians perform over 200 million searches in PubMed. Physicians face challenges while searching PubMed and often miss relevant articles while retrieving too many nonrelevant articles. We investigated whether the use of search filters in PubMed improves searching by physicians.

Impact of PubMed search filters on the retrieval of evidence by physicians

Salimah Z. Shariff PhD, Jessica M. Sontrop PhD, R. Brian Haynes MD PhD, Arthur V. Iansavichus MLIS, K. Ann McKibbon PhD, Nancy L. Wilczynski PhD, Matthew A. Weir MD, Mark R. Speechley PhD, Amardeep Thind MD PhD, Amit X. Garg MD PhD

Competing interests: Several of the authors (Haynes, Wilczynski, McKibbon) have provided PubMed with search filters, some of which are the subject of this study. No other competing interests were declared.

This article has been peer reviewed.

Correspondence to: Dr. Salimah Shariff, salimah.shariff@lhsc.on.ca


Abstract

Background: Physicians face challenges when searching PubMed for research evidence, and they may miss relevant articles while retrieving too many nonrelevant articles. We investigated whether the use of search filters in PubMed improves searching by physicians.

Methods: We asked a random sample of Canadian nephrologists to answer unique clinical questions derived from 100 systematic reviews of renal therapy. Physicians provided the search terms that they would type into PubMed to locate articles to answer these questions. We entered the physician-provided search terms into PubMed and applied two types of search filters alone or in combination: a methods-based filter designed to identify high-quality studies about treatment (clinical queries “therapy”) and a topic-based filter designed to identify studies with renal content. We evaluated the comprehensiveness (proportion of relevant articles found) and efficiency (ratio of relevant to nonrelevant articles) of the filtered and unfiltered searches.

Results: The average physician-provided search terms retrieved 46% of the relevant articles, while 6% of the retrieved articles were nonrelevant (the ratio of relevant to nonrelevant articles was 1:16). The use of both filters together produced a marked improvement in efficiency, resulting in a ratio of relevant to nonrelevant articles of 1:5 (16 percentage point improvement; 99% confidence interval 9% to 22%; p < 0.003) with no substantive change in comprehensiveness (44% of relevant articles found; p = 0.55).

Interpretation: The use of PubMed search filters improves the efficiency of physician searches. Improved search performance may enhance the transfer of research into practice and improve patient care.
Alternatively, this search could be performed with search filters (Figure 1B). Using filters removes the task of generating and including method-specific or topic-specific terms in a search strategy because the filters act as optimized substitutes. For example, applying the nephrology filter eliminates the need to enter renal terms and synonyms in a search query (e.g., chronic kidney disease, end-stage renal disease, chronic renal failure). The nephrology filter, instead, maximizes the retrieval of all renal content (see the nephrology filter strategy in Figure 1B).

In theory, filters should make searching more efficient; however, empiric evidence of this among physicians is lacking. We conducted this study to determine whether the use of methods-based filters and topic-based filters (alone and in combination) improve the efficiency of physician searches in PubMed. The area of renal medicine is an excellent test model because the literature in this field is dispersed across 400 multidisciplinary journals, and many nephrologists search PubMed for information to guide patient care.14,15

Methods

Study design

We previously published a detailed protocol of the methods used in this study.16 Briefly, we identified 100 high-quality systematic reviews about renal therapy using the EvidenceUpdates service (http://plus.mcmaster.ca/EvidenceUpdates/), which prescreens and identifies systematic reviews and meta-analyses that meet strict methodologic criteria and have a high potential for clinical relevance.17,18 All reviews targeted a single clinical question in the area of renal therapy. The selection process is described in Appendix 1 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.101661/-/DC1). The reviews included an average of 16 primary studies (ranging from 2 to 68), collectively covering a wide variety of renal topics (acute renal injury \( n = 24 \), chronic kidney disease \( n = 22 \), dialysis \( n = 22 \), renal transplantation \( n = 20 \), glomerular diseases \( n = 11 \) and other \( n = 1 \)). We transformed the objective statements of the reviews into clinical questions (100 questions total).

Data collection

We surveyed a random sample of Canadian nephrologists about their information-gathering practices (response rate 75%).15 We provided respondents with a unique, randomly selected clinical question focused on therapy that had been generated from the selected systematic reviews. We asked respondents to specify the search terms that they would type into PubMed to retrieve relevant studies to address the clinical question (“search query”). We continued to administer the survey until we received one valid search query for each of the 100 questions.

To test the effect of the filters on physician

<table>
<thead>
<tr>
<th>Clinical Question: What is the efficacy of hepatitis B vaccination in chronic kidney disease patients?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a: Searching without filters</strong></td>
</tr>
<tr>
<td>Search for hepatitis b vaccination chronic kidney disease Go</td>
</tr>
<tr>
<td><strong>b: Searching with filters</strong></td>
</tr>
<tr>
<td>Search for hepatitis b vaccination Go</td>
</tr>
</tbody>
</table>

PubMed Search Translation:

- hepatitis b [MeSH Terms] OR hepatitis b [All Fields]
- vaccination [MeSH Terms] OR "vaccination" [All Fields]
- "kidney failure" [MeSH Terms] OR "kidney failure" [All Fields]
- "chronic kidney disease" [MeSH Terms] OR "chronic kidney disease" [All Fields]
- "acute renal injury" [MeSH Terms] OR "acute renal injury" [All Fields]

Figure 1: PubMed searches without (A) and with (B) filters. This figure was created from the PubMed clinical queries Web interface; this page currently does not feature a “clinical category” section. When we performed searches with the nephrology filter (B), we removed the term “chronic kidney disease” because the filter acts as an optimized substitute for clinical content terms.
searches, we executed each physician-generated search query in PubMed, and then we repeated the search after selecting the appropriate filter(s). When searching with a filter, we removed the terms for study design and/or renal content as appropriate because the filters acted as optimized substitutes. Occasionally respondents provided misspelled search terms, acronyms or other discrepancies. To mimic how a physician might address these discrepancies, a nephrologist, a medical librarian and a computer scientist developed a priori rules by which to modify the physician-generated search queries. All modifications were carried out in duplicate and differences were resolved by consensus (Appendix 2, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.101661/-/DC1). For our primary analysis, we used the modified searches because these improved both the comprehensiveness and efficiency of the nonfiltered searches, therefore producing a more conservative estimate of filter efficacy. This occurred because filters can override misspelled search terms in queries that, if left uncorrected, could result in fewer or no relevant articles being retrieved.

All searches were performed on July 7, 2010. We tested two forms of the clinical queries therapy filter (narrow and broad) and two forms of the nephrology filter (narrow and broad), alone and in combination. The broad filters were designed to maximize the retrieval of relevant articles, whereas the narrow filters were designed to minimize the retrieval of nonrelevant articles. Thus, for each of the 100 clinical questions, we performed nine searches: the original physician-generated searches plus eight filter-aided searches.

For each search result, we calculated the total number of records found, the number of relevant articles found and the position of the relevant articles in the search results. The primary studies included in the original systematic review served as the set of relevant articles (the reference standard).

**Measures and definitions**

Search recall (Table 1) quantifies the comprehensiveness of a search and is defined as the proportion of relevant articles that are retrieved by the search query. Search precision quantifies the efficiency of a search and is defined as the proportion of articles retrieved by the search query that are relevant to the search question (also expressed as the ratio of relevant to nonrelevant articles).

**Statistical analysis**

We used two-sided paired $t$ tests to compare the comprehensiveness and efficiency between the filtered and nonfiltered physician-generated searches. To adjust for multiple comparisons, we used a Bonferroni-corrected $\alpha$ of 0.003 to interpret significance (equivalent to 0.05 divided by 16 comparisons) and report the 99% confidence intervals (CIs).

In our survey, 80% of respondents indicated that they did not scan beyond the first 40 retrieved records, which equates to two default search pages in PubMed. Therefore, we repeated the primary analysis while restricting the PubMed search results to the first 40 records.

Whereas our primary analysis treated each article in the reference set as equally important, some physicians may be more interested in identifying a smaller subset of influential articles. Therefore, we also tested the ability of the filters to retrieve articles deemed to be highly relevant or highly cited. We considered an article to be highly relevant if it was referenced by UpToDate (www.uptodate.com), a popular online service that provides evidence-based summaries for physicians. Of the 100 questions, 79 were adequately covered in UpToDate, and a third of the relevant articles were cited (on average, there were six highly relevant articles per question). We considered an article to be highly cited if its citation count was greater than the median citation count of all relevant articles for that question (on average, there were eight highly cited articles per question).

**Results**

In total, we surveyed 153 nephrologists and received 115 survey responses. Fifteen responses were excluded, leaving a total of 100 search queries for 100 unique clinical questions (eight responses were excluded for a missing or illegible

| Table 1: Formulae for calculating the comprehensiveness and efficiency of Pubmed searches |
|----------------------------------------|-----------------|-----------------|
| Search result | Relevant articles* | Nonrelevant articles |
| Articles found | $a$ | $b$ |
| Articles not found | $c$ | $d$ |
| Comprehensiveness† = $a/(a+c)$ |
| Efficiency‡ = $a/(a+b)$ |

*For each search, relevant articles were defined as the primary studies included in the original systematic review from which the clinical question was derived.
†Measured by search recall (the number of relevant articles found as a proportion of the total number of relevant articles). Also called sensitivity.
‡Measured by search precision (the number of relevant articles found as a proportion of the total number of articles found). Also called positive predictive value.
search query, and seven were excluded because we received a late response after the question had been reassigned to another nephrologist).

The search performance of the physician-generated search queries with and without the use of search filters is summarized in Table 2. On average, the nonfiltered search queries retrieved 46% of the relevant articles (range: 0%–100%). The average efficiency of the nonfiltered search queries was 6% (range: 0%–67%), meaning that the ratio of relevant to nonrelevant articles was 1:16.

Applying either of the nephrology search filters alone produced the greatest improvement in the proportion of relevant articles retrieved (narrow filter increased comprehensiveness by 6.5 percentage points, \( p = 0.003 \)). Applying the narrow form of the clinical queries filter produced the greatest improvement in efficiency (17 percentage points, \( p < 0.003 \)), while comprehensiveness decreased by 8 percentage points (\( p < 0.003 \)). Applying the narrow forms of both the nephrology and clinical queries search filters produced the greatest overall improvement: efficiency improved by 16 percentage points (\( p < 0.003 \)) and comprehensiveness remained relatively unchanged (\( p = 0.55 \)). Expressing this improvement in efficiency another way, the ratio of relevant to nonrelevant articles improved from 1:16 with the nonfiltered searches to 1:5 when both filters were used together (a threefold reduction in nonrelevant articles). No combination of search filters produced simultaneous improvements in both comprehensiveness and efficiency.

When we restricted the results to the first 40 citations (Table 3), the use of the narrow form of the clinical queries filter alone maximally improved overall search performance. Comprehensiveness improved twofold (from 13% to 26%, \( p < 0.003 \)) and efficiency improved fourfold (from 5.5% to 23%, \( p < 0.003 \)), or an improvement in the ratio of relevant to nonrelevant articles from 1:18 in nonfiltered searches to 1:5 for filtered searches. Neither the use of the narrow form of the clinical queries filter nor the addition of the nephrology filters (narrow or broad) substantively affected the improvements achieved.

Filter performance was similar when only highly relevant or highly cited articles were considered (Appendix 3, available at www.cmaj.ca

---

**Table 2: Efficiency and comprehensiveness of physician-generated searches with and without search filters**

<table>
<thead>
<tr>
<th>Clinical queries therapy filter</th>
<th>Nephrology filter</th>
<th>Search performance</th>
<th>Percentage-point change from nonfiltered search (99% CI)</th>
<th>( p ) value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad</td>
<td>Narrow</td>
<td>Broad</td>
<td>Narrow</td>
<td>Efficiency; Comprehensiveness†</td>
</tr>
<tr>
<td>No filter (physician-provided searches)</td>
<td>Efficiency</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td>( \times )</td>
<td></td>
<td></td>
<td></td>
<td>Comprehensiveness</td>
</tr>
</tbody>
</table>

Note: CI = confidence interval.
*Bonferroni-adjusted \( \alpha = 0.003 \).
†Efficiency is the proportion of articles retrieved by the search query that are relevant to the search question (search precision, positive predictive value).
‡Comprehensiveness is the proportion of all relevant articles that are retrieved by the search query (search recall, sensitivity).
The combined use of the narrow forms of both filters produced the greatest overall improvement in search performance with significant improvements in efficiency and little change in comprehensiveness.

**Interpretation**

Our study shows that PubMed filters can improve the efficiency of literature searches conducted by physicians to answer questions about therapy. The combined use of the narrow forms of the clinical queries therapy and nephrology filters resulted in the greatest improvement in search efficiency without reducing the comprehensiveness of the search. For quick searches in which a physician scans fewer than 40 results, the clinical queries therapy filter produced the greatest improvement in both search efficiency and comprehensiveness.

This study moves beyond the development of filters to testing their functionality in the real-world context of physician searching. To our knowledge, the latter has only been attempted twice before and never with methods- and topic-based filters in combination. The two previous evaluations compared the use of the clinical queries filters to standard searches in PubMed and Google Scholar. Although each study found that the use of filters improved efficiency, their conclusions are tempered by the small number of searches performed (range: 3–30) and the limited number of clinical topics tested (3, compared with 100 in our study). The second study also used searches developed by the researchers, which may be less generalizable to searches performed by physicians in busy clinical settings.

The use of the clinical queries filters is strongly recommended by proponents of evidence-based medicine and is often taught in courses for physicians about literature searching. Our results support the use of this PubMed filter. Although we examined questions related to renal medicine, we also achieved excellent results when these methods-based filters were used in the absence of the nephrology filters, which suggests that the results would be generalizable to other subspecialties. However, the use of clinical queries was infrequent among our respondents (20%), despite 89% indicating that they had searched PubMed at least

### Table 3: Efficiency and comprehensiveness of physician-generated searches with and without search filters, limited to the first 40 results retrieved

<table>
<thead>
<tr>
<th>Clinical queries therapy filter</th>
<th>Nephrology filter</th>
<th>Search performance</th>
<th>Percentage-point change from nonfiltered search (99% CI)</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad</td>
<td>Broad</td>
<td>Efficiency‡; Comprehensiveness†</td>
<td>Mean, %</td>
<td></td>
</tr>
<tr>
<td>Narrow</td>
<td>Narrow</td>
<td>Efficiency</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comprehensiveness</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>23.1</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>25.6</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Efficiency</td>
<td>21.7</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong></td>
<td></td>
<td>Comprehensiveness</td>
<td>26.8</td>
<td></td>
</tr>
</tbody>
</table>

Note: CI = confidence interval.
*Bonferroni-adjusted α = 0.003.
†Efficiency is the proportion of articles retrieved by the search query that are relevant to the search question (search precision, positive predictive value).
‡Comprehensiveness is the proportion of all relevant articles that are retrieved by the search query (search recall, sensitivity).
once in the previous year for information to guide patient care.

Strengths and limitations
Our study has several strengths. We followed a published protocol to reduce bias.\textsuperscript{46} To maximize the generalizability of our results to practicing physicians, our survey used the Tailored design method with random selection.\textsuperscript{42} Our survey achieved a 75% response rate, which is considered high for a physician population.\textsuperscript{25,26} We obtained physician-generated searches from a varied group of respondents, including both newer and more seasoned physicians practising in community and academic settings. We used an objective measure of relevance by choosing primary studies included in systematic reviews. Further, we accounted for multiple comparisons by using a conservative level of statistical significance. Finally, our results were consistent across several sensitivity analyses.

We did not directly observe physicians as they searched for medical information, and there may be a discrepancy between what busy physicians report and what they do in practice.\textsuperscript{47} By using primary studies included in systematic reviews, we excluded articles that some physicians find relevant (e.g., studies of lower methodologic quality, narrative reviews, case reports, animal studies, commentaries). However, we used widely accepted principles of the hierarchy of evidence to identify the most important primary articles to retrieve in a search.\textsuperscript{27} Our results proved robust when we restricted our analysis to the most influential articles.

Our study tested the initial physician-generated search query. In practice, a physician may retry an unsuccessful search or modify the search based on the results. Nonetheless, our results indicate that filters can improve the initial search, which may save valuable time by obviating the need for iterative searches. For reasons of feasibility, we focused on questions of therapy as most systematic reviews pertain to prevention and treatment. As more systematic reviews on diagnosis and prognosis are published, similar methods of testing can be used as well.

Implications for further research
While we found that the best performing filter combinations improved efficiency without reducing comprehensiveness, this was not true for all combinations of filters, as some decreased comprehensiveness or efficiency or both. Evaluations are thus necessary to evaluate the real-world effectiveness of the filters and promote knowledge translation. The methods developed for this study can be applied to other disciplines (e.g., cardiology) and by other users (e.g., generalists). Furthermore, although we showed that the combined use of methods- and topic-based filters produced marked improvements in efficiency, over 40%–50% of the highly relevant articles were not retrieved. This highlights that further refinements and innovations in filter development are needed.

Conclusions
The use of search filters can assist physicians to search more effectively, in less time and with less frustration. For quick clinical searches, the clinical queries filters available in PubMed can significantly improve efficiency of physician searching. Improved search performance has the potential to enhance the transfer of research into practice and improve patient care.

References


**Affiliations:** From the Division of Nephrology (Shariff, Iansavichus, Weir, Garg) and the Department of Epidemiology and Biostatistics (Shariff, Sontrop, Speechley, Thind, Garg), University of Western Ontario, London, Ont.; the Department of Medicine (Haynes) and the Department of Clinical Epidemiology and Biostatistics (Haynes, McKibbon, Wilczynski, Garg), McMaster University, Hamilton, Ont.

**Contributors:** Salimah Shariff and Amit Garg had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Salimah Shariff, R. Brian Haynes, Arthur Iansavichus, K. Ann McKibbon, Nancy Wilczynski and Amit Garg contributed to the study concept and design. Salimah Shariff, Arthur Iansavichus and Matthew Weir acquired the data. Salimah Shariff, Jessica Sontrop and Amit Garg performed the statistical analysis. They, along with R. Brian Haynes, K. Ann McKibbon, Mark Speechley and Amardeep Thind interpreted the data. Salimah Shariff and Jessica Sontrop drafted the manuscript. All of the authors critically revised the manuscript for important intellectual content and approved the final version submitted for publication.

**Funding:** This project was supported by the Canadian Institutes of Health Research (CIHR; grant no. 191466). Dr. Shariff was supported by a CIHR Doctoral Research Award and the Schulich Scholarship for Medical Research from the University of Western Ontario. Dr. Thind is supported by a CIHR Canada Research Chair in Health Services Research. Dr. Garg is supported by a Clinician Scientist Award from the CIHR. Dr. Haynes has received grant funding from the Ontario Ministry of Health.

**Acknowledgements:** The authors thank Mr. Dariusz Gozdzik, who programmed the online survey, and Drs. Robert Yang and Arsh Jain, who assisted in assessing the reviews for renal content. The authors also thank Ms. Shayna Bejaimal for assisting with the administration of the survey. This research was conducted in collaboration with the Health Information Research Unit, McMaster University, the Kidney Clinical Research Unit, London Health Sciences Centre, and members of the Department of Epidemiology and Biostatistics, University of Western Ontario.