Dialysis Search Filters for PubMed, Ovid MEDLINE, and Embase Databases

Arthur V. Iansavichus,* R. Brian Haynes,† Christopher W.C. Lee,* Nancy L. Wilczynski,‡ Ann McKibbon,‡ Salimah Z. Shariff,* Peter G. Blake,* Robert M. Lindsay,* and Amit X. Garg*†‡

Summary

Background and objectives Physicians frequently search bibliographic databases, such as MEDLINE via PubMed, for best evidence for patient care. The objective of this study was to develop and test search filters to help physicians efficiently retrieve literature related to dialysis (hemodialysis or peritoneal dialysis) from all other articles indexed in PubMed, Ovid MEDLINE, and Embase.

Design, setting, participants, & measurements A diagnostic test assessment framework was used to develop and test robust dialysis filters. The reference standard was a manual review of the full texts of 22,992 articles from 39 journals to determine whether each article contained dialysis information. Next, 1,623,728 unique search filters were developed, and their ability to retrieve relevant articles was evaluated.

Results The high-performance dialysis filters consisted of up to 65 search terms in combination. These terms included the words “dialy” (truncated), “uremic,” “catheters,” and “renal transplant wait list.” These filters reached peak sensitivities of 98.6% and specificities of 98.5%. The filters’ performance remained robust in an independent validation subset of articles.

Conclusions These empirically derived and validated high-performance search filters should help physicians to effectively retrieve dialysis information from PubMed, Ovid MEDLINE, and Embase.


Introduction

To care for patients, physicians need to access best available evidence. Physicians frequently search bibliographic databases, such as MEDLINE via PubMed, for such information (1,2). Unfortunately, when physicians search PubMed they invariably do not retrieve all relevant articles for the question they are trying to answer, and do retrieve many nonrelevant articles. This stems from the many different ways to describe the same concept. For example, in the area of dialysis, many different terms can be used to describe patients undergoing that treatment. This includes the terms hemodialysis, ESRD, uremic patients, or stage 5 CKD. Similarly, dialysis procedures can be described in many ways, such as renal replacement therapy, hemofiltration, hemodiafiltration, or ultrafiltration. In constructing a search in PubMed, physicians need to navigate variable dialysis-related terminology, inconsistent database indexing, and specific database syntax.

One way to resolve this problem is to use an information search filter. Take, for example, a clinician searching for information on how vascular access screening for fistulas and grafts improves clinically relevant outcomes in hemodialysis patients (Figure 1A).

As presented in Figure 1A, a clinician could type in terms “(fistulae OR graft) AND surveillance AND hemodialysis.” The alternative would be not to type any dialysis-related terms at all but rather simply apply a dialysis search filter that comprehensively identifies all the dialysis-related articles available in PubMed. As presented in Figure 1B, a clinician would then simply type in the words “(fistulae OR graft) AND surveillance,” ignoring all the other related dialysis terms.

The benefit of this would be that clinicians would no longer be searching the entire database and would not have to type in dialysis-related terms. Rather, with their two remaining terms they would be searching within a select number of articles in PubMed known to be dialysis relevant. Similar in concept to disease screening in a higher-risk population, this would reduce the number of nonrelevant articles that are retrieved while increasing the chance of finding relevant articles. To further develop this concept, we sought to develop and test search filters to help physicians efficiently retrieve articles with dialysis information (hemodialysis or peritoneal dialysis) indexed in PubMed, Ovid MEDLINE, and Embase.

Materials and Methods

Study Overview

We used a diagnostic test analytic framework with separate development and validation sets of articles to develop and test various search filters. We first
manually reviewed the full texts of 22,992 articles to determine whether they contained dialysis information. This formed our reference standard. We developed a computer program to test each term separately in PubMed, MEDLINE, Ovid, and Embase using different syntax and search fields; download the results; and store them for further analysis. We then used a computer-generated algorithm to combine single terms into complex search filters and compared the ability of all developed filters to retrieve articles with dialysis information. For each filter, we constructed a 2×2 contingency table (Supplemental Table A) and evaluated its performance by calculating sensitivity, specificity, precision, and accuracy. Filters with the best operating characteristics in the development set of articles were then tested in the independent validation set of articles. This process was repeated separately for filters designed for PubMed, Ovid MEDLINE, and Embase. Key characteristics of each database are presented in Supplemental Table B.

Sample of Articles

We applied an article sampling strategy similar to that used in previously published studies because this has resulted in search filters that generalize well over publication years and journal types (3–5). We first compiled a list of 466 journals that had published at least one article relevant to renal care. We then selected 20 high-yield journals and 19 journals randomly sampled from the remaining 446 journals. These 39 journals were randomly divided into development and validation subsets of articles using a ratio of 1 to 2 (Supplemental Table C). We then manually reviewed the full texts of 22,992 articles published in these 39 journals from 2004 to 2008 (articles indexed in PubMed, Ovid MEDLINE, and Embase). We reviewed all types of articles published in these journals, including original investigations, reviews, case reports, letters, and editorials.

Review of Each Article for Relevance to Dialysis

We developed a checklist to determine whether an article contained dialysis-relevant information (Supplemental Table D). All forms of short- and long-term dialysis were considered eligible, including hemodialysis, peritoneal dialysis, and continuous veno-venous hemofiltration. All reviewers applied this checklist to determine whether each full-text article was relevant to dialysis. Using two sets of 100 articles, all the reviewers were trained and calibrated in their use of checklist criteria (agreement beyond chance, κ = 0.91).

Filters

We compiled renal dialysis- and various nephrology-related text words and index terms from the following sources: U.S. National Library of Medicine Medical Subject Heading (MeSH) thesaurus (6), MEDLINE permuted index (7), EMTREE medical thesaurus (8), SNOMED medical terminology database (version 2006.2.0030) (9), nephrology textbooks (10,11), clinical practice guidelines (12,13), 195 renal systematic reviews (14), online nephrology glossaries, and clinicians and health sciences librarians from eight different countries. All selected terms were presented as single terms in both American and British English spelling (e.g., hemodialysis and haemodialysis), truncated and nontruncated terms (truncation symbol “*” allows searching multiple endings of a term, such as “hemodi*”), abbreviated terms (e.g., ESRD, CAPD, CCPD, CVVH), as well as selected phrases (e.g., renal hyperparathyroidism). Examples of search terms considered for the filters were “dialysis,” “hemodialysis,” “end-stage renal disease,” “uremia,” and “renal replacement therapy.” These terms were considered for all search fields available in each bibliographic database, such as record title, abstract, part of indexing term, or related text words and index terms from the following tax and terminology. With use of a previously described algorithm, single terms were combined into multiple term filters (3).

Statistical Analyses

We calculated the sensitivity, specificity, precision, and accuracy of each filter (formulas are presented in Supplemental Table A). Filters with the best operating characteristics were then further evaluated in the independent validation subset of articles.
Results

Reference Standard

We reviewed the full texts of 22,992 articles published in 39 journals, and 5.1% of them contained information relevant to dialysis (Supplemental Table C). A total of 14,619 articles from 26 journals formed the development set, and 8073 articles from the remaining 13 journals formed the validation set.

Sample Terms and Single-Term Filters

We used a total of 97,221 unique search terms from all sources, which we then used to generate 1,623,728 unique search filters. The single-term filters with the optimal performance were truncated terms “dialy*[tw],” with 79.4% sensitivity and 98.7% specificity for PubMed, and “dialy$.mp,” with 79.2% sensitivity and 98.7% specificity for Ovid MEDLINE. The exploded EMTREE term “renal replacement therapy” had 78.9% sensitivity and 98.5% specificity for Embase.

Multiple-Term Filters

We devised 1,362,473 multiple-term search filters and evaluated their performance in the development set of articles. Our best-performing high-sensitivity and high-specificity dialysis filters for PubMed, Ovid MEDLINE, and Embase databases are presented in Table 1. These filters used up to 65 terms in combination and included such search terms as “dialysis,” “hemodialysis,” “end stage renal disease,” “uremia,” “renal replacement therapy,” “hemofiltration,” “hemodiafiltration,” “ultrafiltration,” “artificial kidney,” and “renal transplant wait list.” These filters reached peak sensitivities of 98.6% and specificities of 98.5%. The filters’ performance remained robust in the independent validation set of articles.

Proof-of-Concept Searches

To analyze the utility of our PubMed high-sensitivity and high-specificity filters, in an exploratory analysis we asked five nephrologists to provide a search query that they would use to answer a focused clinical question in the area of dialysis. Each question provided to a nephrologist was unique and was derived from a recently published systematic review (wherein the primary articles cited in the review represented the relevant articles of interest [15–19]). Compared with unaided searches, the retrieval characteristics of a nephrologist’s nondialysis terms improved when our dialysis filters were used. Both highly sensitive and highly specific dialysis content filters increased the yield of relevant articles (Table 2). But this was frequently at the expense of retrieving additional nonrelevant articles, particularly with the high-sensitivity filter.

Discussion

People perceive that dialysis-related articles can be easily identified in bibliographic databases. Unfortunately, this is often not true. Frequently, only full-text examination of an article reveals that its content pertains to patients receiving dialysis. The article abstract or index terms may describe such patients as simply having CKD or ESRD, or having renal osteodystrophy or left ventricular hypertrophy in the setting of kidney failure. Without a more robust strategy, such articles relevant to the care of dialysis patients would be missed in clinicians’ searches. On the other hand, some articles mention dialysis patients, but only as an exclusion criterion (i.e., patients receiving maintenance dialysis were excluded from trial participation) (20–22). Other articles may mention that patients are not yet undergoing dialysis. Such articles and other nonrelevant articles are often retrieved with some simple nonoptimized dialysis free-text search terms but can be avoided when more complex multiterm searches are used. In this study we succeeded in developing and validating high-performance dialysis filters for busy clinicians and researchers.

Using Dialysis Filters

Our best dialysis filters are balanced for high sensitivity (which may be more suitable for academics working to identify all relevant articles for practice guidelines or systematic reviews) or for high specificity (to meet the needs of busy clinicians who are looking for a quick answer and who don’t have time to sift through many irrelevant citations). As of January 2012, our most sensitive dialysis filter reduced the number of records retrieved from PubMed from over 21 million to about 210,000; the most specific filter reduced this number to about 120,000 records.

Physicians can use these filters in many ways. These filters can be directly incorporated into the database search engine, as was done in PubMed for filters developed by members of our team to identify articles with high methodologic rigor, such as clinical trials (23). These filters can also be typed in (or copied and pasted in) while a database is directly searched, then saved there for future use. In the meantime, we provide access to these filters developed for PubMed at http://hiru.mcmaster.ca/hiru/hiru_hedges_nephrology_.filters.aspx. On this page, anyone can enter into the search field any nondialysis search terms for their topic of interest. Users then select the “Dialysis” filter and can decide whether they want to use a broad or narrow form of the filter. For the information needs of many physicians, a high-specificity (narrow) filter will be sufficient. Physicians can use these filters on their own or in combination with other available filters to further refine the retrieved results. For example, physicians interested only in retrieving systematic reviews for a topic of interest could simply select use of the “Systematic Reviews” filter on this page as well.

Strengths and Limitations

To our knowledge, this is the first rigorous attempt to develop search filters in the area of dialysis. This complements our work done to develop information filters for kidney transplantation (24) and general nephrology (3). Better access to dialysis information is important because dialysis patients have high morbidity and mortality, with high costs to health care systems. Use of these filters is an efficient way of handling the large amount of dialysis information scattered across millions of citations in large bibliographic databases and could pay important dividends for health professionals, their patients, and society. In this study we developed and tested more than 1.5 million unique filters for PubMed, Ovid MEDLINE, and Embase interfaces. The optimized high-performance filters reached sensitivities and specificities in excess of 98%.

However, the efficiency of these filters is subject to some limitations. The actual performance of searches performed by
Table 1. Optimized high-sensitivity and high-specificity dialysis filters for PubMed, Ovid MEDLINE and Embase

<table>
<thead>
<tr>
<th>Filters</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
<th>Precision % (95% CI)</th>
<th>Accuracy % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High-Sensitivity Filter for PubMed</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation Set: 97.3 (95.3–99.3)</td>
<td>95.9 (95.5–96.4)</td>
<td>44.2 (40.2–48.3)</td>
<td>96.0 (95.5–96.4)</td>
<td></td>
</tr>
<tr>
<td><strong>High-Specificity Filter for PubMed</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation Set: 92.3 (89.1–95.5)</td>
<td>98.4 (98.2–98.7)</td>
<td>66.3 (61.4–71.2)</td>
<td>98.2 (98.0–98.5)</td>
<td></td>
</tr>
<tr>
<td>Filters</td>
<td>Sensitivity % (95% CI)</td>
<td>Specificity % (95% CI)</td>
<td>Precision % (95% CI)</td>
<td>Accuracy % (95% CI)</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>High-Sensitivity Filter for Ovid MEDLINE</strong></td>
<td>Development Set: 96.4 (95.1–97.7)</td>
<td>95.2 (94.8–95.6)</td>
<td>56.4 (53.8–59.1)</td>
<td>95.3 (94.9–95.6)</td>
</tr>
<tr>
<td></td>
<td>Validation Set: 97.3 (95.3–99.3)</td>
<td>96.0 (95.6–96.4)</td>
<td>44.8 (40.7–48.9)</td>
<td>96.1 (95.6–96.5)</td>
</tr>
<tr>
<td><strong>High-Specificity Filter for Ovid MEDLINE</strong></td>
<td>Development Set: 90.1 (88.9–92.9)</td>
<td>98.5 (98.3–98.7)</td>
<td>80.0 (77.4–82.6)</td>
<td>98.1 (97.8–98.3)</td>
</tr>
<tr>
<td></td>
<td>Validation Set: 92.7 (89.5–95.8)</td>
<td>98.5 (98.2–98.7)</td>
<td>66.5 (61.6–71.4)</td>
<td>98.3 (98.0–98.5)</td>
</tr>
</tbody>
</table>
### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Filters</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
<th>Precision % (95% CI)</th>
<th>Accuracy % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High-Sensitivity Filter for Ovid EMBASE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h?emodia$.mp OR dialy$.mp OR exp renal replacement therapy/ OR ((end stage or endstage) adj (kidney OR renal)).tw OR ur?emia.mp OR renal replacement.mp OR esrd.tw OR renal osteodystroph$.mp OR (autosomal dominant polycystic kidney).ti OR artificial kidney / OR renal hyperparathyroidism.tw OR ((&quot;kidney failure&quot; / OR (chronic adj (kidney or renal))).tw OR &quot;chronic kidney disease&quot; / OR ckd.tw) AND (erythropoietin.mp OR c reactive protein / OR calcification.mp OR vascular.ti OR hypertrophy.mp OR &quot;hypertension&quot; / OR nephrogenic.mp OR stage.tw)) OR (((kidney or renal) adj transplant$).tw AND (candidates.tw OR wait$ list$.tw)) OR (((indwelling catheter/ OR exp vein catheterization/) AND (exp kidney failure/ OR heparin.mp)) OR (ur?emic.tw OR hyppar$.tw OR hyperpar$.tw) AND (kidney.ti OR renal.ti OR ckd.ti)) OR (arteriovenous fistula/ AND (catheter$.mp OR nephrolog$.tw)))</td>
<td>Development Set: 98.6 (97.8–99.4)</td>
<td>95.2 (94.8–95.6)</td>
<td>57.4 (54.8–59.9)</td>
<td>95.4 (95.1–95.8)</td>
</tr>
<tr>
<td>Validation Set: 98.3 (96.8–99.8)</td>
<td>96.2 (95.8–96.6)</td>
<td>48.7 (44.7–52.8)</td>
<td>96.3 (95.9–96.7)</td>
<td></td>
</tr>
<tr>
<td><strong>High-Specificity Filter for Ovid EMBASE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(exp *renal replacement therapy/ OR h?emodialy$.tw OR peritoneal dialysis.mp OR dialy$.ti OR h?emodialysis patient$.mp OR dialysis patient$.tw OR exp *dialysis/ OR ((end stage or endstage) adj (kidney or renal)).ti OR dialysis therapy.tw OR sevelamer.mp OR artificial kidney / OR h?emodialift$.mp OR esrd$ti OR renal replacement.ti OR h?emofti$.tw OR intradialy$.tw OR renal hyperparathyroidism.tw OR nephrogenic systemic fibrosis.tw OR ur?emic patient$.tw OR ((renal replacement.mp OR ((kidney or renal) adj failure)).tw OR ur?emia.tw OR catheter$.mp OR erythropoietin$.mp OR fistula $.mp OR &quot;chronic kidney failure&quot; / OR hyperkalemia.mp OR hyperoxaluria.tw OR exp blood vessel calculation/ OR anuria.mp) AND (dialys$.mp OR h?emodialys$.mp)) OR (((chronic adj (kidney or renal)).ti OR ur?emi$.ti OR ckd.ti) AND (inflammation.tw OR renal osteodystrophy.mp OR hypertrophy.tw)) OR (((kidney or renal) adj transplant$.tw AND exp *urinary tract disease/ AND (candidates.tw OR wait list$.tw)))</td>
<td>Development Set: 93.4 (91.7–95.0)</td>
<td>98.5 (98.3–98.7)</td>
<td>80.3 (77.9–82.8)</td>
<td>98.2 (98.0–98.4)</td>
</tr>
<tr>
<td>Validation Set: 92.8 (89.8–95.7)</td>
<td>98.6 (98.4–98.9)</td>
<td>71.2 (66.6–75.7)</td>
<td>98.4 (98.2–98.7)</td>
<td></td>
</tr>
</tbody>
</table>


b Ovid Medline fields: $= truncation character; mp=multiple posting (term appears in title, abstract, or MeSH); tw=term present in title; /=MeSH character; adj= adjacent operator; *=focused MeSH term; adj2= defined adjacency operator; exp=exploded MeSH term; ?=optional wildcard; ti=term present in title

c Ovid Embase fields: exp=exploded EMTREE term; /=EMTREE character; adj= adjacent operator; $= truncation character; mp=multiple posting (term appears in title, abstract, or EMTREE); tw=term present in title or abstract; *=focused EMTREE term; adj2= defined adjacency operator; ec=Embase section headings field; ti=term present in title
<table>
<thead>
<tr>
<th>Clinical Question</th>
<th>Relevant Articles Retrieved (n)</th>
<th>Nonrelevant Articles Retrieved (n) *</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physician</td>
<td>Physician Search with Best Sensitive Dialysis Content Filter</td>
<td>Physician Search with Best Specific Dialysis Content Filter</td>
</tr>
<tr>
<td>What bleeding risk is associated with long-term antiplatelet use in the end stage renal disease population? (14 relevant articles)</td>
<td>6</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Does vascular access screening for fistulas and grafts improve clinically relevant outcomes in hemodialysis patients? (10 relevant articles)</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>What is the efficacy of L-carnitine supplementation for treatment of patients with intradialytic hypotension and cramping? (6 relevant articles)</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>What is the risk of homocysteine on total mortality and cardiovascular disease for patients with end-stage renal disease? (27 relevant articles)</td>
<td>17</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>What is the impact of antiplatelet agents on the thrombosis rates of arteriovenous fistulae and grafts used for hemodialysis access? (10 relevant articles)</td>
<td>4</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

*The number of nonrelevant articles retrieved is expressed per relevant article and was rounded up to the nearest whole number. The search phrases as determined and typed in by physicians were: “(bleeding OR blood) AND (platelet OR antiplatelet) AND end stage renal disease,” “hemodialysis AND surveillance AND (fistulae OR graft),” “intradialytic AND L-carnitine,” “(homocysteine AND end-stage renal disease,” “thrombosis AND platelet AND hemodialysis access.” Renal dialysis terms were removed when the physician search was combined with dialysis search filter.
physicians with or without dialysis filters will also depend on quality of other, nondialysis terms entered by a physician. When poorly chosen, misspelled, or incorrectly combined, these terms could greatly hinder overall retrieval performance. Also, one may find a lower level of precision (also called a positive predictive value) from the values reported in Table 1 when these filters are applied to the entire 21 million articles available via PubMed. However, the reported sensitivity and specificity of each filter will not change substantially. Furthermore, some articles that are incorrectly indexed in a particular database will be difficult or impossible to find no matter how rigorous the search. Physicians who require a comprehensive search should search several bibliographic sources. Finally, many information gaps exist in areas where no research for dialysis patients has been done, and one cannot expect to retrieve information when it does not exist in a bibliographic database.

Practice Implications and Future Research

Evaluation of published evidence is a critical component in evidence-based practice. Health professionals often do not have enough time or expertise to comprehensively search for evidence. The high-performance filters we have developed here represent a better method of efficient and comprehensive searching for dialysis information available through major bibliographic databases. Future research is now needed to measure the effect of different methods of information retrieval on physicians’ knowledge, treatment choices, clinical decision-making, and patient care.

Acknowledgments

We thank Mr. Chris Cotoi and Mr. Nicholas Hobson for their help and support with programming and data support.

This project was supported by an operating grant from the Canadian Institutes of Health Research (CIHR). C.W.C.L. was supported by a Schulich Research Opportunities Program Award. A.X. G. was supported by a CIHR Clinician-Scientist Award.

Disclosure

None.

References

7. U.S. National Library of Medicine: Permutated Medical Subject Headings, Bethesda, MD, National Library of Medicine, 2003

Received: March 5, 2012 Accepted: July 9, 2012

Published online ahead of print. Publication date available at www.cjasn.org.

This article contains supplemental material online at http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.02360312/-/DCSupplemental.