AJKD
Original Investigation

# High-Performance Information Search Filters for CKD Content in PubMed, Ovid MEDLINE, and EMBASE

Arthur V. Iansavichus, MLIS,[1] Ainslie M. Hildebrand, MD,[1,2] R. Brian Haynes, MD, PhD,[3] Nancy L. Wilczynski, PhD,[3] Adeera Levin, MD,[4] Brenda R. Hemmelgarn, MD, PhD,[5] Karen Tu, MD, MSc,[6,7,8] Gihad E. Nesrallah, MD, MSc,[9,10] Danielle M. Nash, MSc,[1,3] and Amit X. Garg, MD, PhD[1,2,3,11]

**Background:** Finding relevant articles in large bibliographic databases such as PubMed, Ovid MEDLINE, and EMBASE to inform care and future research is challenging. Articles relevant to chronic kidney disease (CKD) are particularly difficult to find because they are often published under different terminology and are found across a wide range of journal types.

**Study Design:** We used computer automation within a diagnostic test assessment framework to develop and validate information search filters to identify CKD articles in large bibliographic databases.

**Setting & Participants:** 22,992 full-text articles in PubMed, Ovid MEDLINE, or EMBASE.

**Index Test:** 1,374,148 unique search filters.

**Reference Test:** We established the reference standard of article relevance to CKD by manual review of all full-text articles using prespecified criteria to determine whether each article contained CKD content or not. We then assessed filter performance by calculating sensitivity, specificity, and positive predictive value for the retrieval of CKD articles. Filters with high sensitivity and specificity for the identification of CKD articles in the development phase (two-thirds of the sample) were then retested in the validation phase (remaining one-third of the sample).

**Results:** We developed and validated high-performance CKD search filters for each bibliographic database. Filters optimized for sensitivity reached at least 99% sensitivity, and filters optimized for specificity reached at least 97% specificity. The filters were complex; for example, one PubMed filter included more than 89 terms used in combination, including "chronic kidney disease," "renal insufficiency," and "renal fibrosis." In proof-of-concept searches, physicians found more articles relevant to the topic of CKD with the use of these filters.

**Limitations:** As knowledge of the pathogenesis of CKD grows and definitions change, these filters will need to be updated to incorporate new terminology used to index relevant articles.

**Conclusions:** PubMed, Ovid MEDLINE, and EMBASE can be filtered reliably for articles relevant to CKD. These high-performance information filters are now available online and can be used to better identify CKD content in large bibliographic databases.

*Am J Kidney Dis.* 65(1):26-32. 

***INDEX WORDS:*** Chronic kidney disease (CKD); health informatics; information retrieval; MEDLINE; EMBASE; bibliographic database; search filter; renal insufficiency.

**Editorial, p. 1**

As the amount of medical research published globally continues to grow, in any given subject area it is becoming increasingly difficult to identify articles to guide patient care and inform future research. Most of the information on medical research is stored in online bibliographic databases such PubMed, Ovid MEDLINE, and EMBASE. These databases now house millions of records; for example, PubMed alone contains more than 23 million records with 20,000 records added every week.[1,2] Although more than 2.5 billion PubMed searches were performed in 2013, users frequently failed to retrieve articles relevant to their search.[2-6]

*From the [1]Kidney Clinical Research Unit, London Health Sciences Centre; [2]Division of Nephrology, Western University, London; [3]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario; [4]Division of Nephrology, University of British Columbia, Vancouver, British Columbia; [5]Division of Nephrology, University of Calgary, Calgary, Alberta; [6]Department of Family and Community Medicine, University of Toronto; [7]Toronto Western Hospital Family Health Team, University Health Network; [8]Institute for Clinical Evaluative Sciences; [9]Division of Nephrology, Humber Regional Hospital; [10]The Li Ka Shing Knowledge Institute, Keenan Research Centre, St. Michael's Hospital, Toronto; and [11]Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada.*

*Received March 29, 2014. Accepted in revised form June 2, 2014. Originally published online July 21, 2014.*

*Address correspondence to Ainslie M. Hildebrand, MD, Kidney Clinical Research Unit, Room ELL-112, London Health Sciences Centre, 800 Commissioners Road East, London, Ontario, Canada N6A 4G5. E-mail: ainslie.hildebrand@londonhospitals.ca*



*0272-6386/$36.00*
*http://dx.doi.org/10.1053/j.ajkd.2014.06.010*

Chronic kidney disease (CKD) is a serious and costly condition, with an estimated prevalence of 8%-16% worldwide.[7] However, CKD articles are particularly difficult to find, even for the experienced searcher. Articles relevant to CKD are featured in a broad range of journals and are often indexed inconsistently, with variable terminology used to describe the condition.[8,9] For example, when authors refer to patients with a gradual loss of kidney function as having "renal insufficiency" or "progressive renal disease," the article gets indexed as such and may not be retrieved by using the phrase "chronic kidney disease" in the search strategy. Similarly, articles may be indexed by only clinical or histologic diagnoses, such as "diabetic nephropathy," "glomerulosclerosis," or "kidney fibrosis," making them difficult to capture even with broad search strategies. This, along with variations in indexing and syntax, make searching for CKD-related information in bibliographic databases complicated.

A solution to this challenge is to use search filters designed specifically to capture information relevant to CKD. A CKD search filter would act as an optimized substitute for the variable terminology used to describe this condition. For example, users searching for the benefits and harms of using aldosterone antagonists in CKD may choose to enter the terms "aldosterone antagonists AND chronic kidney disease" in an unaided search query. With this strategy, relevant articles may be missed due to variations in the way the concept of CKD is expressed and indexed in the database. Use of a CKD information search filter would allow users to simplify search terms (eg, to "aldosterone antagonists" alone) while restricting the retrieved search results to CKD-related content. In this case, use of the search filter increases the positive predictive value of the search, analogous to the increase in positive predictive value achieved when a disease screening test is applied to a high-risk population. More relevant articles and fewer nonrelevant articles are retrieved with the remaining search terms.

In this report, we describe how we used computer automation to develop high-performance information search filters to identify articles relevant to CKD in PubMed, Ovid MEDLINE, and EMBASE bibliographic databases. These high-performance CKD search filters proved valid when searching a separate set of articles, and we illustrate how searches were improved when the filter was used in some proof-of-concept physician searches.

## METHODS

### Study Overview

We used a diagnostic test assessment framework to develop and validate search filters for CKD. For the purpose of assessment, CKD content was defined by a study population with CKD stage 3, 4, or 5; any study that examined chronic changes in kidney function or proteinuria; or any study that elucidated basic science processes related to CKD (Table S1, available as online supplementary material).[10]

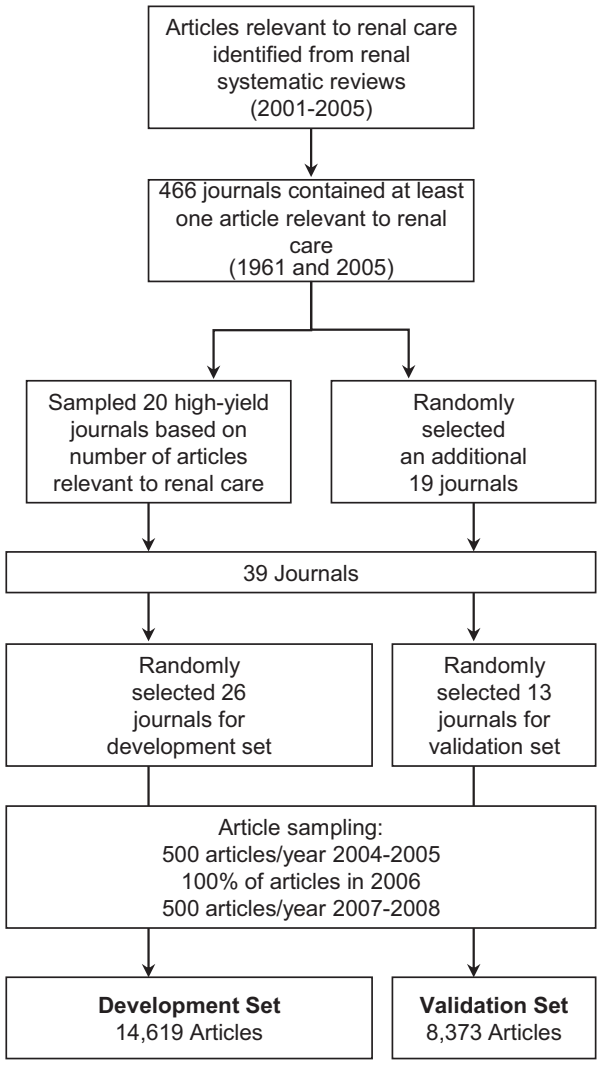


**Figure 1.** Methods for creation of development and validation sets of articles.

### Sample of Articles

We first established the reference standard of article relevance to CKD by manual review of a subset of full-text articles published in 39 journals between 2004 and 2008. To develop this collection of journals, we adopted a similar strategy for sampling as published in prior search filter studies (Fig 1).[11] We first assembled a set of 466 journals that had published at least one article relevant to renal care between 1961 and 2005. We then sampled the top 20 high-yield journals (those with the greatest number of articles relevant to renal care) and randomly selected 19 additional journals from the remaining 446. We randomly divided this set of 39 journals into development and validation subsets using a ratio of 2:1 (Table S2). For each journal in the development and validation set, we manually reviewed all full-text articles (including original investigations, reviews, letters, and editorials) indexed in PubMed, Ovid MEDLINE, and EMBASE in 2006 and randomly selected an additional 500 full-text articles per year for the remaining years