Published in final edited form as:

J Clin Epidemiol. 2009 September ; 62(9): 974–981. doi:10.1016/j.jclinepi.2008.11.006.

Diagnostic Test Systematic Reviews: Bibliographic search filters ("Clinical Queries") for diagnostic accuracy studies perform well

M Kastner^{a,b}, NL Wilczynski^a, KA McKibbin^a, AX Garg^{a,c,d}, and RB Haynes^{a,e,*}

^aHealth Information Research Unit, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

^bDepartment of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Ontario, M5T 3M6, Canada

^cDivision of Nephrology, University of Western Ontario, London, Ontario, Canada N6A 4G5

^dDepartment of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, N6A 4G5, Canada

^eDepartment of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

Abstract

Background—Systematic reviews of health care topics are valuable summaries of all pertinent studies on focused questions. However, finding all relevant primary studies for systematic reviews remains challenging.

Objectives—To determine the performance of the Clinical Queries (CQ) sensitive search filter for diagnostic accuracy studies for retrieving studies for systematic reviews.

Methods—We compared the yield of the sensitive CQ diagnosis search filter for MEDLINE and EMBASE to retrieve studies in diagnostic accuracy systematic reviews (*ACP Journal Club*, 2006).

Results—12 of 22 diagnostic accuracy reviews (452 included studies) met inclusion criteria. After excluding 11 studies not in MEDLINE or EMBASE, 95% of articles (417/441) were captured by the sensitive CQ diagnosis search filter (MEDLINE and EMBASE combined). Of 24 studies not retrieved by the filter, 22 were not diagnostic accuracy studies. Re-analysis of the CQ filter without these 22 non-diagnosis articles increased its performance to 99% (417/419). We found no substantive impact of the 2 articles missed by the CQ filter on the conclusions of the systematic reviews in which they were cited.

Conclusion—The sensitive CQ diagnostic search filter captured 99% of articles and 100% of substantive articles indexed in MEDLINE and EMBASE in diagnostic accuracy systematic reviews.

What is new?

© 2009 Elsevier Inc. All rights reserved.

*Corresponding author: RB Haynes, Department of Clinical Epidemiology and Biostatistics, McMaster Health Sciences Centre, McMaster University, 1200 Main Street West, Room 3V43c, Hamilton, Ontario, Canada L8N 3Z5. Tel.: +1-905-525-9140 ext. 20153; fax: +1-905-546-0401. bhaynes@mcmaster.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

- The empirically derived clinical queries (CQ) diagnosis search filter performed well for retrieving articles for diagnostic accuracy systematic reviews.
- The CQ diagnosis search filter is a useful and readily available tool when commencing searching for original articles to conduct a systematic review.
- Clinicians and researchers conducting diagnostic accuracy systematic reviews can begin with the sensitive CQ diagnosis search filter in MEDLINE and EMBASE. Additional strategies will be needed for other databases. Future studies should compare this strategy with others for performance.

BACKGROUND

Systematic reviews are valuable resources for clinicians and researchers because they summarize all pertinent studies on a specific clinical question, can improve the understanding of inconsistencies among diverse evidence, help users to keep up with the medical literature, define future research agendas, and inform the management of health problems^{1, 2}. However, finding all primary studies for systematic reviews is challenging because an overwhelming amount of information is available in the biomedical literature. In addition, complete, accurate retrieval is compromised by indexing inconsistencies and ambiguities, and lack of empirically validated searching filters (also referred to as search strategies and hedges)³.

One approach is to use a complex search filter based on the principles of library science, such as that of the InterTASC Information Specialists' Sub-Group⁴. A potential alternative is to use the simple but sensitive search filters originally designed to assist clinicians to find the current best evidence for clinical decisions. Our group developed empirically validated search filters for several purposes including retrieving higher quality studies of treatment⁵, diagnosis⁶, etiology⁷, prognosis⁸, and clinical prediction guides⁹. These are publicly available in the Clinical Queries (CQ) interface of MEDLINE

(<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>) as well as the limits screen of Ovid¹⁰ for MEDLINE, EMBASE, PsycINFO, and CINAHL. Three types of CQ search filters are available in Ovid: “sensitive” (retrieves a high proportion of relevant or on-target articles, but also a suboptimal number of off-target articles reflect the low precision figures), “specific” (somewhat lower sensitivity but fewer off-target retrievals), and “optimal” (best balance of sensitivity and specificity); the optimal strategy is not available on the CQ page in PubMed.

In this investigation, we sought to determine how well the MEDLINE and EMBASE sensitive CQ diagnostic search filters retrieved the diagnostic accuracy studies included in systematic reviews. We looked at this question in 2 ways. First, we determined if the included studies were retrievable by using CQ filters. Second, we calculated if the use of CQ filters reduced the number of potentially relevant studies that needed to be reviewed after searching in MEDLINE and EMBASE.

METHODS

The methods used to derive the CQ search filters in both MEDLINE and EMBASE have been described elsewhere^{5–9}. In this study, we compared the yield of the sensitive CQ diagnostic filters (Table 1) with the studies included in a sample of systematic reviews of diagnostic accuracy from the *ACP Journal Club* collection (<http://www.acpjc.org>) for the year 2006 (search done in May, 2007).

Diagnostic accuracy reviews were searched from the *ACP Journal Club* web site by entering “Review” in the search field, and selecting “Diagnosis” from the “Article type” drop-down

menu. We selected diagnostic accuracy reviews by first looking at each *ACP Journal Club* page of titles for reviews that were bannered as 2006. Study eligibility was determined by looking at each 2006 diagnosis review in full-text, which was downloaded using the PubMed identifying numbers hyperlink at the end of the citation of the original study. Eligibility criteria for including a diagnostic accuracy systematic review in our study were that it was published in 2006, incorporated a MEDLINE and EMBASE search as a data source, and that the review was available and downloadable in electronic format. When the diagnostic accuracy reviews were analyzed in full-text, we discovered that the systematic review by Wardlaw et al¹¹ used our diagnostic CQ search filter as part of their strategy. To avoid “incorporation bias”, we added this additional criterion for eligibility: the systematic review could not use the CQ search filters. The included studies of each of the eligible diagnostic systematic reviews were documented in an Excel datasheet.

For each eligible diagnostic review topic, we ran the sensitive CQ diagnostic search filter in both MEDLINE and EMBASE using the Ovid Technologies interface. Starting with MEDLINE, each of the included studies for each systematic review was located by entering citation information in the search field. Once an included study was located, the “diagnosis (sensitivity)” option in the “Limits” tab was used to test if the article would be captured by the sensitive CQ diagnostic search filter.

We assessed the effect on the conclusions of a review of any included study that was not retrieved by the sensitive search filter. We defined an included study as having a “potential impact” on the summary measures and conclusions of the systematic review if it was used as one of studies included in a meta-analysis, or if it was described in the Results section of the review in the context of any of the outcomes. For articles included in the review but not retrieved by the sensitive search filter, we defined a non-retrieved article as having “no impact” on the systematic review if the exclusion of the included study from the review’s analysis made no difference to the final conclusions compared with results if the study had been included in the analysis.

To determine if the sensitive CQ diagnostic search filter reduces the number of studies that need to be screened after searching MEDLINE and EMBASE, we sought to replicate the search filters in the systematic reviews. We contacted the authors of the systematic reviews to obtain the exact filter used in their search.

RESULTS

Of 94 diagnosis accuracy reviews found in *ACP Journal Club*, 22 were published in 2006. When reviewed in full-text, 13 systematic reviews met our original inclusion criteria (both MEDLINE and EMBASE searches as data sources, and available in electronic format)^{11–23}. The addition of the third eligibility criterion during full-text review resulted in the exclusion of 1 diagnostic review¹¹; and 9 systematic reviews were excluded because they did not meet our original inclusion criteria: 1 systematic review did not explicitly include a MEDLINE search filter, 6 reviews did not include an EMBASE search filter, and 2 reviews were not available in electronic format, leaving 12 systematic reviews for our sample^{12–23}. A total of 452 studies were analyzed by the 12 systematic reviews (Table 2). Of these, 11 articles from 2 reviews^{12, 13} were abstracts from conference proceedings and were not indexed in either MEDLINE or EMBASE, and thus were excluded from the analysis (Figure 1).

Performance of the sensitive Clinical Query diagnostic search filter

Figure 1 shows the flow diagram of the process that was used to calculate the proportion of articles that were captured by the most sensitive CQ diagnosis search filter. After excluding the 11 abstracts not indexed in MEDLINE or EMBASE from the pool of 452 included studies,

95% of articles (417/441) were captured by the sensitive CQ search filter when results from MEDLINE and EMBASE were combined. Of these, 273 articles (62%) overlapped between MEDLINE and EMBASE, 114 articles (26%) were captured in MEDLINE but not EMBASE, and 30 articles (7%) were found in EMBASE only.

The 24 articles (5%) that were missed by the search strategy were from 6 systematic reviews (Table 3)^{14–17, 22, 23}. We explored the characteristics of the missed articles to determine why they were missed by our sensitive filter. We reviewed the titles and abstracts of all 24 missed articles to determine if they were about diagnostic accuracy. We used a simplified version of criteria previously developed by Wilczynski et al²⁴ to determine this: the study compared at least 2 diagnostic test procedures with one another. Of the 24 non-retrieved articles, 22 articles did not meet the criteria of a diagnosis study (Table 3). Treatment accounted for 16 (66.7%), 4 (16.7%) were classified as ‘something else’ (defined as: content of the study does not fit any of the definitions for other purpose categories [e.g., diagnosis, treatment]²⁴), 1 (4.2%) was an etiologic study²⁵, and 1 (4.2%) was classified as a prognosis study because even though patients with dyspepsia underwent 2 tests (whole blood serology and endoscopy) to determine the frequency of gastroesophageal cancer²⁶ the investigation did not include a direct comparison of these diagnostic tests for detecting gastroesophageal cancer.

We then re-calculated the proportion of articles that the sensitive CQ diagnostic search filter retrieved by excluding these 22 non-diagnosis articles from the sample, giving a rate of 99% (417/419). The missing 1% represents 2 articles^{27, 28}, from separate systematic reviews^{14, 15}, that were missed by the sensitive search filter. This proportion of articles was consistent with the performance characteristics of the sensitive diagnostic search filter in MEDLINE (sensitivity 98.6%, specificity 74.3%)⁶ and EMBASE (sensitivity 100%, specificity 70.4%)²⁹.

We assessed the impact of these 2 articles by excluding their findings from their reviews to see if this would affect the conclusions of the reviews. The first missed diagnosis article (Kiilholma et al²⁷) was one of 121 included studies of a systematic review by Martin et al¹⁴, which compared ≥ 2 diagnostic techniques with a gold standard (multichannel urodynamics) for diagnosing urinary incontinence. We found that the removal of this study from the systematic review had no impact on the results because the study was not used in any pooled analyses or described in the Results section for the outcomes.

The second non-retrieved diagnosis study (Dahele et al²⁸) was one of 34 included studies of a systematic review that compared the performance of the endomysial antibody (EMA) test with two types of tTG antibody tests (i.e., human recombinant [hr] and guinea pig [gp]) to make recommendations for the most appropriate screening test for celiac disease. The sensitivities and specificities of the 34 studies were pooled in a meta-analysis. We re-calculated the meta-analysis of the included studies to determine if the removal of the study by Dahele et al would impact on the overall results (using Meta-DiSc, version 1.4). We found an absolute increase of 0.3% in the pooled sensitivity for the tTG-Ab diagnostic test when the study by Dahele et al was removed, but no difference was found between the two sets of pooled specificities (Table 4). A similar absolute increase (0.2%) was found for the pooled sensitivities of the EMA diagnostic test when the study by Dahele et al was removed, and no difference between the 2 sets of pooled specificities (Table 4). Because this change was so small and within the span of the 95% CIs (see Table 4), we can conclude that the review was not substantively affected by the exclusion of this single study.

We reviewed these 2 articles, both of which include diagnostic accuracy data, to determine the reasons they were not retrieved by the sensitive CQ diagnostic search filters. Dahele et al²⁸ included no terms or phrases related to diagnosis or diagnostic accuracy testing in the title,

abstract, or indexing terms in MEDLINE and it is not indexed in EMBASE. Our request to the National Library of Medicine to have the MEDLINE indexing reviewed resulted in re-indexing. Now this article is retrieved by the sensitive CQ diagnostic filter. The article by Kiilholma et al has no diagnostic information in its EMBASE record. Its MEDLINE record includes the subheading ultrasonography, defined as the use of ultrasonography in the diagnosis of diseases. This ultrasonography subheading is not included in the sensitive CQ diagnostic filter.

We next sought to replicate the search filters in the articles to determine if the use of the sensitive CQ diagnostic search filter may have saved time during screening of studies after searching. We received a response from 7 authors (58%), but only 5 provided a detailed search filter for MEDLINE^{12, 13, 18, 21, 22}. We included a 6th systematic review in our analysis because it provided the exact search filter used in the Ovid Technologies interface within the manuscript¹⁴. Using the Ovid interface in MEDLINE, we reproduced the search filters of these 6 systematic reviews by entering all search terms within the publication date parameters provided. Before testing, 3 of the 6 search filters required consensus from 3 investigators to clarify their interpretation of the search filter with respect to Boolean operator placement, and translation between PubMed and Ovid interfaces. After the searches were entered we applied the sensitive CQ diagnostic search filter to each search yield. Overall we found that 5 of the 6 systematic reviews showed a 35% to 63% reduction in the number of articles that would have to be assessed for relevance compared with retrievals without the use of the MEDLINE sensitive CQ diagnostic search filter. In the 6th systematic review, the number of articles retrieved remained the same after applying the sensitive CQ diagnostic search filter.

DISCUSSION

We showed that the sensitive CQ diagnostic search filter performed extremely well by capturing 99% of included articles indexed in MEDLINE and EMBASE from our sample of 12 diagnostic systematic reviews. The original capture estimate of 95% was improved once we determined that 92% of the 24 missed articles (22/24) were not about diagnosis. A large proportion of missed articles that were not about diagnosis occurred in reviews that addressed both diagnosis and treatment questions. Other explanations for non-retrieved articles include inconsistencies in indexing. One of the 2 non-retrieved articles about diagnosis was missed because of incomplete indexing, and has now been re-indexed. The second article missed by the sensitive CQ filter can be attributed to the performance of the operating characteristics of the filter itself, which does not include the diagnostic subheading ultrasonography. Further, the single missed diagnostic test study that was part of a pooled analysis did not have a substantive effect on the review conclusion.

Others have investigated the usefulness of search filters to identify diagnostic accuracy studies^{30, 31}. Both of these studies concluded that search filters do not perform well enough to warrant their use in finding articles for systematic reviews. However, these studies differ in a number of ways from ours. First, neither examined the non-retrieved articles to determine whether they met the definition of diagnostic accuracy studies (including a comparison of at least 2 diagnostic tests) or whether these studies would have materially affected the conclusions of the reviews in which they appeared. Second, both confined their searching to MEDLINE. Restricting our analysis to articles indexed in Medline, we observed 89% retrieval for our sensitive filter, without adjustment for studies that did not meet the definition of diagnostic accuracy: this is consistent with the findings of Leeftang et al (87% retrieval for our sensitive filter)³⁰. Third, neither considered ways that the search filter results could be extended, for example by examining references in the appropriately retrieved studies. (We did not consider this either, as our retrieval was very high; however, it seems that examining references in retrieved studies would be a logical extension of their findings that could have tempered their

negative conclusions.) Ritchie et al³¹ also included only 1 systematic review and used content terms that may have limited the retrieval of the search filters they tested. Leeftang et al³⁰ also reported considerable variability in results from study to study which could explain the differences in findings of the 3 investigations.

Strengths and limitations

Our study aimed to evaluate the performance of an empirically-derived search filter as a tool for retrieving articles for a diagnostic accuracy systematic review. We used empirically validated search filters for both MEDLINE and EMBASE and examined in detail the original studies that were not retrieved by these filters, determining that most of the non-retrieved studies did not report comparisons of 2 or more tests, and that 2 studies that did report such comparisons did not substantively alter the conclusions of the reviews in which they were reported. Furthermore, the confirmation that incorrect indexing was the reason that 1 of these 2 studies were not retrieved by our diagnostic search filter, strengthens the interpretation of our findings. The search filters we used are readily available and can be used on Ovid, PubMed and EBSCO: CINAHL interfaces. Our results show promise for clinicians and researchers conducting systematic reviews because the sensitive search filter is easily used and retrieved all key articles in MEDLINE and EMBASE. Additional searches would be needed for other databases and for studies reported only in abstract form in conference proceedings.

As for any search filter in MEDLINE or EMBASE, the retrieval of the sensitive filter includes many off-target (“false positive”) articles that will need to be assessed and eliminated. We set out to test the efficiency of the sensitive CQ diagnostic search filter by comparing it with the search filters used in the 12 systematic reviews, but only one review provided the exact search strategy used¹⁴, and only 5 authors provided a detailed strategy for MEDLINE on request. We were unable to replicate the searches exactly, partly because of errors in the search filters, or the lack of details provided concerning Boolean operator placement, or uncertainties about translation for the Ovid interface. This finding is consistent with previous reports that indicate that errors in search filters are frequently revealed when the strategy is provided in enough detail to attempt replication³². This also raises an important question—to what extent can systematic reviews be replicated?

Our study has some additional limitations. First, our findings are based on a sample of systematic reviews from *ACP Journal Club*. We chose *ACP Journal Club* as the source for reviews because it includes only higher quality systematic reviews from a limited, but important set of clinical general healthcare and specialist journals. Second, we looked only in MEDLINE and EMBASE for primary articles included in the reviews; we did not look for articles that were not indexed in MEDLINE or EMBASE, but might be found in other databases (such as BIOSIS, PsycINFO, and so on) or the grey literature, which are considered important in the systematic review process. However, over 97% of the articles in the reviews we studied were indexed by MEDLINE, EMBASE or both. The Cochrane Handbook for Diagnostic Test Accuracy Reviews does not recommend the use of search filters because “studies examining the accuracy of diagnostic tests are not well indexed in the electronic bibliographic databases, such as MEDLINE”³³. Our search filter would appear to circumvent this limitation. However, we are not suggesting that use of our sensitive filter would obviate the need for additional searching for studies as part of the process of conducting a systematic review of diagnostic test accuracy studies. Rather, we believe that it is a useful tool for beginning such searching and, if verified in other studies, many suffice for the MEDLINE and EMBASE database searches. It should be noted that the appropriate content terms (typically terms related to the clinical problem or diagnostic test of interest) must be added to the CQ filters. Further studies are needed to assess the reproducibility of searches reported in systematic reviews, the relative

precision of search filters, and the performance of search filters in other databases and for other purposes, such as reviews of therapy, prognosis, and etiology.

REFERENCES

1. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376–380. [PubMed: 9054282]
2. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. Reviewing the reviewers: the quality of reporting in three secondary journals. *CMAJ* 2001;164:1573–1576. [PubMed: 11402795]
3. Ely JW, Osheroff JA, Ebell MH, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ* 2002;324:710. [PubMed: 11909789]
4. InterTASC Information Specialists' Sub-Group Search Filter Resource. 2008 [Accessed January 23]. Available at: <http://www.york.ac.uk/inst/crd/intertasc/diag.htm>
5. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre S. for the Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytic survey. *BMJ* 2005;330:1179–1182. [PubMed: 15894554]
6. Haynes RB, Wilczynski NC. for the Hedges Team. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytic survey. *BMJ* 2004;328:1040–1042. [PubMed: 15073027]
7. Wilczynski NL, Haynes RB. for the Hedges Team. Developing optimal search strategies for detecting clinically sound causation studies in MEDLINE. *Proc AMIA Annu Symp* 2003;719–723.
8. Wilczynski NL, Haynes RB. for the Hedges Team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE. *BMC Medicine* 2004;2:23. [PubMed: 15189561]
9. Wong, SS.; Wilczynski, NL.; Haynes, RB.; Ramkissoonsingh, R. Hedges Team. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE; *AMIA Annu Symp Proc*; 2003. p. 728-732.
10. Clinical queries in Ovid. 2008 [Accessed on July 9]. Available at:
=B86UCj8j&p_accessibility=0&p_redirect=&p_lva=&p_sp=cF9zcmNoPTEmcF9zb3J0X2J5PSZwX2dyaWRzb3J0PSZwX3Jvd19jbnQ9MTUsMTUmcF9wcm9kc20wJnBfY2F0cz0wJnBfcHY9JnBfY3Y9JnBfc2VhcmNoX3R5cGU9YW5zd2Vycy5zZWYy2hfbmwmcF9wYWdlPTEmcF9zZWYy2hfdGV4dD1jbGluaWNhbCBxdWVyaWVz&p_li=&p_topview=1.
[http://ovidsupport.custhelp.com/cgi-bin/ovidsupport.cfg/php/enduser/std_adp.php?](http://ovidsupport.custhelp.com/cgi-bin/ovidsupport.cfg/php/enduser/std_adp.php?p_faqid=1599&p_created=1087487498&p_sid)
[p_faqid=1599&p_created=1087487498&p_sid](http://ovidsupport.custhelp.com/cgi-bin/ovidsupport.cfg/php/enduser/std_adp.php?p_faqid=1599&p_created=1087487498&p_sid)
11. Wardlaw JM, Chappell FM, Best JJK, Warolowska K, Berry E. Non-invasive imaging compared with intra-arterial angiography in the diagnosis of symptomatic carotid stenosis: a meta-analysis. *Lancet* 2006;367:1503–1512. [PubMed: 16679163]
12. Gisbert JP, de la Morena F, Abaira V. Accuracy of Monoclonal Stool Antigen Test for the diagnosis of *H. pylori* infection: A systematic review and meta-analysis. *Am J Gastroenterol* 2006;101:1921–1930. [PubMed: 16780557]
13. Davenport C, Cheng EYL, Kwok YT, et al. Assessing the diagnostic test accuracy of natriuretic peptides and ECG in the diagnosis of left ventricular systolic dysfunction: a systematic review and meta-analysis. *Br J Gen Practice* 2006;56:48–56.
14. Martin JL, Williams KS, Abrams KR, et al. Systematic review and evaluation of methods of assessing urinary incontinence. *Health Technol Assess* 2006;10:1–132. [PubMed: 16487456]iii–iv
15. Lewis NR, Scott BB. Systematic review: the use of serology to exclude or diagnose celiac disease (a comparison of the endomysial and tissue transglutaminase antibody tests). *Aliment Pharmacol Ther* 2006;24:47–54. [PubMed: 16803602]
16. Vakil N, Moayyedi P, Fennerty MB, Talley NJ. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: Systematic review and meta-analysis. *Gastroenterol* 2006;131:390–401.
17. Avouac J, Gossec L, Dougados M. Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review. *Ann Rheum Dis* 2006;65:845–851. [PubMed: 16606649]

18. Battaglia M, Pewsner D, Juni P, Egger M, Bucher HC, Bachmann LM. Accuracy of B-Type natriuretic peptide tests to exclude congestive heart failure. *Arch Intern Med* 2006;166:1073–1080. [PubMed: 16717169]
19. Colli A, Fraquelly M, Casazza G, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: A systematic review. *Am J of Gastroenterol* 2006;101:513–523. [PubMed: 16542288]
20. de Graaf, Prak A, Bierma-Zeinstra S, Thomas S, Peuo W, Koes B. Diagnosis of lumbar spinal stenosis: A systematic review of the accuracy of diagnostic tests. *Spine* 2006;31:1168–1176. [PubMed: 16648755]
21. Gisbert JP, Abairra V. Accuracy of *Helicobacter pylori* diagnostic tests in patients with bleeding peptic ulcer: A systematic review and meta-analysis. *Am J Gastroenterol* 2006;101:848–863. [PubMed: 16494583]
22. Miyasaki JM, Shannon K, Voon V, et al. Practice Parameter: Evaluation and treatment of depression, psychosis, and dementia in Parkinson disease (an evidence-based review): Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 2006;66:996–1002. [PubMed: 16606910]
23. Suchowersky O, Reich S, Perlmuter J, Zesiewicz T, Gronseth G, Weiner WJ. Practice Parameter: Diagnosis and prognosis of new onset Parkinson disease (an evidence-based review): Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 2006;66:968–975. [PubMed: 16606907]
24. Wilczynski NL, Morgan D, Haynes RB. and the Hedges Team. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak* 2005;5:20. [PubMed: 15969765]
25. Locascio JJ, Corkin S, Grwodon JH. Relation between clinical characteristics of Parkinson's disease and cognitive decline. *J Clin Exp Neuropsychol* 2003;25:94–109. [PubMed: 12607175]
26. Sung JJY, Lao WC, Lai MS, et al. Incidence of gastroesophageal malignancy in patients with dyspepsia in Hong Kong: implications for screening strategies. *Gastrointest Endosc* 2001;54:454–458. [PubMed: 11577306]
27. Kiilholma PJ, Makinen JI, Pitkanen YA, et al. Perineal ultrasound: an alternative for radiography for evaluating stress urinary incontinence in females. *Ann Chir Gynaecol* 1994;208:43–45.
28. Dahele AVM, Aldhous MC, Humpreys K, et al. Serum IgA tissue transglutaminase antibodies in coeliac disease and other gastrointestinal diseases. *QJM* 2001;94:195–205. [PubMed: 11294962]
29. Wilczynski NL, Haynes RB. and the Hedges Team. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Med* 2005;3:7. [PubMed: 15796772]
30. Leeftang MMG, Scholten RJPM, Rutjes AWS, Reitsma JB, Bossuyt PMM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol* 2006;59:234–240. [PubMed: 16488353]
31. Ritchie G, Glanville J, Lefebvre C. Do published search filters to identify diagnostic test accuracy studies perform adequately? *Health Inform Libr J* 2007;24:188–192.
32. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol* 2006;59:1057–1063. [PubMed: 16980145]
33. The Cochrane Handbook for Diagnostic Test Accuracy Reviews. 2008 [Accessed on July 1 Available at]. Available at <http://srdta.cochrane.org/en/authors.html>

Figure 1.

Flow diagram of the process for analyzing the proportion of articles that were captured by the best sensitivity Clinical Query (CQ) diagnosis search filters

Table 1

Best sensitivity Clinical Queries for studies of diagnostic accuracy in MEDLINE and EMBASE

Search interface	Search filter in MEDLINE	Sensitivity*
Ovid	sensitiv:.mp. OR diagnos:.mp. OR di.fs.	98.6%
PubMed	(sensitiv* [Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnos* [Title/Abstract] OR diagnosis[MeSH:noexp] OR diagnostic [MeSH:noexp] OR diagnosis,differential[MeSH:noexp] OR diagnosis[Subheading:noexp])	
Search interface	Search filter in EMBASE	Sensitivity
Ovid	di.fs. OR predict:.tw. OR specificity.tw.	100%

* Sensitivity = proportion of high quality articles retrieved.

Table 2

Characteristics of 12 diagnostic systematic reviews published in 2006*

Systematic review	Journal	Objective	Discipline	Number of included studies
Avouac J, et al ¹⁷	Annals of Rheumatic Diseases	To evaluate the two generations of anti-cyclic citrullinated protein (CCP) antibodies as a diagnostic marker of rheumatoid arthritis (RA) and as a predictor of future development of RA	Rheumatology (RA)	65
Battaglia M, et al ¹⁸	Archives of Internal Medicine	To compare the diagnostic accuracy of BNP tests (ELISA and RIA assays) for diagnosing congestive heart failure	Cardiology	19
Colli A, et al ¹⁹	American Journal of Gastroenterology	To compare the accuracy of US, spiral CT, MRI and AFP in diagnosing HCC	GI (HCC)	30
Davenport C, et al ¹³	British Journal of General Practice	To investigate the accuracy of the ECG, BNP, NT-proBNP and the combinations of two or more tests in the diagnosis of LVSD in primary care.	Cardiology (primary care)	32
de Graaf I, et al ²⁰	Spine	To evaluate the diagnostic value of imaging, clinical and other tests used to detect lumbar spinal stenosis	Orthopedics	24
Gisbert JP and Abaira V ²¹	American Journal of Gastroenterology	To compare the accuracy of the different tests aimed to detect <i>H. pylori</i> infection in patients with upper GI bleeding	GI (upper GI bleeding)	23
Gisbert JP and de la Morena F, et al ¹²	American Journal of Gastroenterology	To perform a systematic review of the accuracy of monoclonal SAT both for the initial diagnosis of <i>H. pylori</i> infection and for the confirmation of its eradication after treatment	GI (<i>H. pylori</i>)	26
Lewis NR, et al ¹⁵	Alimentary Pharmacology & Therapeutics	To compare the sensitivities and specificities of the enomysial antibody and the tissue transglutaminase antibody tests	GI (celiac disease)	34
Martin JL, et al ¹⁴	Health Technology Assessment	To identify and synthesize studies of diagnostic processes of UI and to construct an economic model to examine the cost-effectiveness of simple, commonly used primary care tests	Urology (UI)	120
Miyasaki JM ²²	Neurology	To make evidence-based treatment recommendations for patients with Parkinson disease (PD), dementia, depression, and psychosis	Neurology (PD)	22
Suchowersky O, et al ²³	Neurology	To define key issues in the diagnosis of Parkinson disease (PD), to define features influencing progression, and to make evidence-based recommendations	Neurology (PD)	39
Vakil N, et al ¹⁶	Gastroenterology	To determine the diagnostic accuracy of alarm features in predicting malignancy by performing a meta-analysis based on the published literature	GI (upper GI malignancy)	18

* Systematic review = a narrowly focused synthesis of the literature on a particular topic with replicable methodology, and which may or may not include a meta-analysis; ELISA = enzyme-linked immunosorbent assay; RIA = radioimmunosorbent assay; US = ultrasound scan; CT = computed tomography; MRI = magnetic resonance imaging; AFP = alpha-fetoprotein; HCC = hepatocellular carcinoma; ECG = electrocardiogram; BNP = brain natriuretic peptide; NT-proBNP = N terminal-pro BNP; LVSD = left ventricular systolic dysfunction; *H. pylori* = helicobacter pylori; SAT = stool antigen test.

Table 3
Included studies of systematic reviews which were not captured by the sensitive CQ diagnostic search filter

Systematic review and its included studies	Purpose category
Miyasaki (n = 11)	
Aarsland D, Laake K, Larsen JP, Janvin C. Donepezil for cognitive impairment in Parkinson's disease: a randomised controlled study. <i>J Neurol Neurosurg Psychiatry</i> 2002;72:708–712.	Treatment
Andersen J, Aabro E, Gulman N, Helmsted A, Pedersen H. Antidepressive treatment in Parkinson's disease. A controlled trial of the effect of nortriptyline in patients with Parkinson's disease treated with L-Dopa. <i>Acta Neurol Scand</i> 1980;62:210–219.	Treatment
Avila A, Cardona X, Martin Baranera M, Maho P, Satre F, Bello J. Does nefazodone improve both depression and Parkinson disease? A pilot randomized trial. <i>J Clin Psychopharmacol</i> 2003;23:509–513.	Treatment
Breier A, Sutton VK, Feldman PD, et al. Olanzapine in the treatment of dopaminergic-induced psychosis in patients with Parkinson's disease. <i>Biol Psychiatry</i> 2002;52:438–445.	Treatment
Fregni F, Santos CM, et al. Repetitive transcranial magnetic stimulation is as effective as fluoxetine in the treatment of depression in patients with Parkinson's disease. <i>J Neurol Neurosurg Psychiatry</i> 2004;75:1171–1174.	Treatment
Leentjens AF, Vreeling FW, Luijckx GJ, Verhey FR. SSRIs in the treatment of depression in Parkinson's disease. <i>Int J Geriatric Psychiatry</i> 2003;18:552–554.	Treatment
Morgante L, Epifanio A, Spina E, et al. Quetiapine versus clozapine: a preliminary report of comparative effects on dopaminergic psychosis in patients with Parkinson's disease. <i>Neurol Sci</i> 2002;23:S89–90.	Treatment
Neufeld MY, Blumen S, Aitkin I, Parmet Y, Korczyn AD. EEG frequency analysis in demented and nondemented parkinsonian patients. <i>Dementia</i> 1994;5:23–28.	Something else
Parkinson Study Group. Low-dose clozapine for the treatment of drug-induced psychosis in Parkinson's disease. <i>N Engl J Med</i> 1999;340:757–763.	Treatment
Rektorova I, Rektor I, Bares M, et al. Pramipexole and pergolide in the treatment of depression in Parkinson's disease: a national multicentre prospective randomized study. <i>Eur J Neurol</i> 2003;10:399–406.	Treatment
Wermuth L, Sorensen P, Timm S, et al. Depression in idiopathic Parkinson's disease treated with citalopram. <i>Nord J Psychiatry</i> 1998;52:163–169.	Treatment
Martin (n = 5)	
Berglund AL, Lalos O. The pre- and postsurgical nursing of women with stress incontinence. <i>J Adv Nurs</i> 1996;23:502–11.	Treatment
Kiilholma PJ, Makinen JI, Pitkanen YA, Varpula MJ. Perineal ultrasound: an alternative for radiography for evaluating stress urinary incontinence in females. <i>Ann Chir Gynaecol Suppl</i> 1994;208:43–5.	Diagnosis
Robinson D, McClish DK, Wyman JF, Bump RC, Fanti JA. Comparison between urinary diaries completed with and without intensive patient instructions. <i>Neurourol Urodyn</i> 1996;15:143–8	Treatment
Sandvik H, Seim A, Vanvik A, Hunskaar S. A severity index for epidemiological surveys of female urinary incontinence: comparison with 48-hour pad-weighing tests. <i>Neurourol Urodyn</i> 2000;19:137–45.	Something else
Williams KS, Assassa RP, Smith NKG, Jagger C, Perry S, Shaw C, et al. Development, implementation and evaluation of a new nurse-led continence service: a pilot study. <i>Journal of Clinical Nursing</i> 2000;9:566–73.	Treatment
Avouac (n = 4)	
Bobbio-Pallavicini F, Alpini C, Caporali R, Avasse S, Bugatti S, Montecucco C. Autoantibody profile in rheumatoid arthritis during long-term infliximab treatment. <i>Arthritis Res Ther</i> 2004;6:R264–72.	Treatment
Bogliolo L, Alpini C, Caporali R, Scire CA, Moratti R, Montecucco C. Antibodies to cyclic citrullinated peptides in psoriatic arthritis. <i>J Rheumatol</i> 2005;32:511–15.	Something else
Caramaschi P, Biasi D, Tonolli E, Pieropan S, Martinelli N, Carletto A, et al. Antibodies against cyclic citrullinated peptides in patients affected by rheumatoid arthritis before and after infliximab treatment. <i>Rheumatol Int</i> 2005;26:58–62.	Treatment
Chen HA, Lin KC, Chen CH, Liao HT, Wang HP, Chang HN, et al. The effect of etanercept on anti-cyclic citrullinated peptide antibodies and rheumatoid factor in patients with rheumatoid arthritis. <i>Ann Rheum Dis</i> 2006;65:35–9.	Treatment
Suchowersky (n = 2)	
Locascio JJ, Corkin S, Growdon JH. Relation between clinical characteristics of Parkinson's disease and cognitive decline. <i>J Clin Exp Neuropsychol</i> 2003;25:94–109.	Etiology
Thaisethawatkul P, Boeve BF, Benarroch EE, et al. Autonomic dysfunction in dementia with Lewy bodies. <i>Neurology</i> 2004;62:1804–1809.	Something else
Lewis (n = 1)	

Systematic review and its included studies		Purpose category
Dahele AVM, Aldhous MC, Humpreys K, et al. Serum IgA tissue transglutaminase antibodies in coeliac disease and other gastrointestinal diseases. Q J Med 2001;94: 195–205.		Diagnosis
Vakil (n = 1)		
Sung JJY, Lao WC, Lai MS, Li TH, Chan FKL, Wu JCY, Leung VKS, Luk YW, Kung NNS, Ching JYL, Leung WK, Lau J, Chung SJY. Incidence of gastroesophageal malignancy in patients with dyspepsia in Hong Kong: implications for screening strategies. Gastrointest Endosc 2001;54:454–458.		Prognosis
Proportion of articles that are about diagnosis		2/24 = 8%

Something else = content of the study does not fit any of the definitions for other purpose categories such as treatment, etiology, prognosis, diagnosis, clinical prediction rule according to Haynes et al^{5–9}

Results of meta-analyses of included studies from the diagnostic systematic review by Lewis et al with or without the study by Dahele et al*

Pooled operating characteristic (95% CI) [†]	Diagnostic test					
	tTG-Ab			EMA		
	All studies	Excluding Dahele et al	Absolute difference (CI)	All studies	Excluding Dahele et al	Absolute difference
Sensitivity	93.4% (92.5 to 94.1)	93.7% (92.9 to 94.5)	0.3% (0.4 to 0.4)	93.5% (92.6 to 94.3)	93.7% (92.8 to 94.5)	0.2% (0.2 to 0.2)
Specificity	96.3% (95.9 to 96.6)	96.3% (95.9 to 96.6)	0	99.6% (99.5 to 99.7)	99.6% (99.5 to 99.7)	0

* CI = confidence interval; tTG-Ab = tissue transglutaminase antibody; EMA = endomysial antibody.

[†] Meta-analyses were done using Meta-Disc, version 1.4.