Optimal search strategies for detecting health services research studies in MEDLINE

Nancy L. Wilczynski, R. Brian Haynes, John N. Lavis, Ravi Ramkissoonsingh, Alexandra E. Arnold-Oatley; for the HSR Hedges Team

Abstract

Background: Evidence from health services research (HSR) is currently thinly spread through many journals, making it difficult for health services researchers, managers and policy-makers to find research on clinical practice guidelines and the appropriateness, process, outcomes, cost and economics of health care services. We undertook to develop and test search terms to retrieve from the MEDLINE database HSR articles meeting minimum quality standards.

Methods: The retrieval performance of 7445 methodologic search terms and phrases in MEDLINE (the test) were compared with a hand search of the literature (the gold standard) for each issue of 68 journal titles for the year 2000 (a total of 25 936 articles). We determined sensitivity, specificity and precision (the positive predictive value) of the MEDLINE search strategies.

Results: A majority of the articles that were classified as outcome assessment, but fewer than half of those in the other categories, were considered methodologically acceptable (no methodologic criteria were applied for cost studies). Combining individual search terms to maximize sensitivity, while keeping specificity at 50% or more, led to sensitivities in the range of 88.1% to 100% for several categories (specificities ranged from 52.9% to 97.4%). When terms were combined to maximize specificity while keeping sensitivity at 50% or more, specificities of 88.8% to 99.8% were achieved. When terms were combined to maximize sensitivity and specificity while minimizing the differences between the 2 measurements, most strategies for HSR categories achieved sensitivity and specificity of at least 80%.

Interpretation: Sensitive and specific search strategies were validated for retrieval of HSR literature from MEDLINE. These strategies have been made available for public use by the US National Library of Medicine at www.nlm.nih.gov/nichsr /hedges/search.html.

Wﬕith the increasing emphasis on "using evidence" and "value for money" in health services, it is essential that researchers, clinicians, health system managers and public policy-makers be able to retrieve relevant, high-quality reports of health services research (HSR). Efficiently retrieved research evidence can aid in decision-making about which services to provide and in the resource allocation decisions to support those services, reducing the need for arbitrary decisions and aiding collaboration with clinicians and consumers.[1] MEDLINE is a huge and ex-

panding bibliographic resource that is freely available to all with Internet access. Yet the volume of the literature often overwhelms both clinicians and health system decision-makers.[2,3] End-users of MEDLINE and other large bibliographic databases have difficulty executing precise searches[2,3] and are often unaware of what kind of information to seek, where to find it[3,4] and how to judge its quality.[3]

HSR has been defined as the scientific study of the effect of health care delivery; the organization and management of health care access, quality, cost and financing; and the evaluation of the impact of health services and technology (Allmang NA, Koonce TY. Health services research topic searches. Bethesda [MD]: National Library of Medicine; 2000. Unpublished report). More recently, HSR has been defined as the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies and personal behaviours affect access to health care, the quality and cost of health care and, ultimately, health and well-being.[5] HSR articles constitute only a tiny fraction of the MEDLINE database and are spread through a large number of journals; hence, MEDLINE searching is challenging. Conversely, journal browsing is impractical as a means of retrieving all relevant studies for a given question or staying abreast of the literature. Our aim was to develop methodologic search filters for MEDLINE to enable end-users to efficiently retrieve articles of relevance to clinical practice guidelines (CPGs) and the appropriateness, process, outcomes, cost and economics of health services.

Methods

We compared the retrieval performance of methodologic search terms and phrases in MEDLINE with a manual review of each article in each issue of 68 journal titles for the year 2000 for the study categories of appropriateness, process assessment, outcome assessment, CPGs, cost and economics of care.

Search terms

Candidate content and methodologic terms (text words and Medical Subject Headings [MeSH] [exploded and nonexploded], publication types) were compiled by reviewing "gold standard" articles and their MEDLINE indexing, the definitions in Table 1 and the criteria in Table 2; by consulting experts in bibliographic database searching for HSR topics (mainly health sciences librari-

ans); and by consulting experts in studying HSR-related questions. All suggested search terms were tested. The terms are available on request from the corresponding author.

Table 1: Definitions of health services research (HSR) categories

HSR category	Definition
Appropriateness	Content pertained directly to determining whether people who received a health care service had the appropriate clinical indications to receive that service
Process assessment	Content pertained directly to assessing the process of care for people with a given health problem, such as who did what to whom, why, where, when and how well
Outcome assessment	Content pertained directly to the appraisal of various clinical decisions and management paths and their effects on patient well-being (outcomes)
Clinical practice guidelines	Content pertained directly to the development or application of statements intended to assist practitioner or patient decisions about appropriate health care (for specific clinical circumstances)
Cost	Content pertained directly to the costs or financing of a health care problem or service
Economics	Content pertained directly to a comparison of the costs and effects of at least 2 different forms of service provision

McMaster HSR database

A database of journals containing relevant studies of HSR was created. We looked for journals that published adequate numbers of relevant articles, such that manual searching of these journals would provide an adequate benchmark against which the MEDLINE searches could be compared. Three independently derived lists were examined to identify appropriate journals.

The first list comprised journals that are reviewed by 4 publications: *ACP Journal Club*, *Evidence-Based Medicine*, *Evidence-Based Nursing* and *Evidence-Based Mental Health*. These publications provide synopses of the articles in 170 journals with the intent of giving health care workers an overview of new developments in medicine and nursing; the journals have been selected on the basis of their yield of studies that meet explicit criteria for methodologic merit and relevance to clinical practice.[11] This list of 170 journals was reduced to 161 by including only those that were indexed in MEDLINE and by conducting hand searches of issues for the year 2000 to determine which journals had published at least 1 study concerning the appropriateness, process or outcomes of care or CPGs.

The second set of journals was derived from a survey by Elixhauser and associates[12] of HSR literature for studies of the economics of health care and the Science Citation Index listing of top-rated journals in the field for the category health care sciences and services. We deleted 2 pharmacy journals from this list because we judged them too narrow in focus for our purposes. Input by a convenience sample of policy-makers led to 2 journal nominations and resulted in 10 unique HSR journals (i.e., 10 titles that were not included in the first list). The third list consisted of journals identified in a report on MEDLINE searches for HSR written by 2 Na-

Table 2: Criteria used to determine the methodologic rigour of journal articles in each HSR category

HSR category*	Methodologic criteria
Appropriateness[6]	Article included a statement of the explicit criteria for appropriateness of care that were applied.
	Data source was independent of the study investigators, or there was an assessment of the reliability of the application of the criteria and the auditors were blinded to practitioner identity and, if more than 1 institution was involved, institution identity.
Process assessment[6]	Article included a statement of the explicit criteria for process of care that were applied.
	Data source was independent of the study investigators, or there was an assessment of the reliability of the application of the criteria and the auditors were blinded to practitioner identity and, if more than 1 institution was involved, institution identity.
Outcome assessment[7]	At least 1 of the outcomes was objective or derived from a data source that was independent of the study.
Clinical practice guidelines[8]	
Development	Article included an explicit statement describing the process for developing the guidelines, including methods of evidence assembly, method of review of studies and at least 1 of the following: organizations and individuals involved, methods of formulating guidelines, and methods of reaching agreement or consensus.
	Evidence was cited in support of at least 1 of the recommendations.
Application	At least 1 of the exact guidelines was provided in a table, figure or the text of the study.
Cost	None.
Economics[9]	Question was a comparison of alternatives.
	Alternative services or activities were compared in terms of outcomes produced (effectiveness) and resources consumed (costs).
	Evidence of effectiveness was from a study of real patients that met criteria (described elsewhere[10]) for diagnosis, treatment, quality improvement or a systematic review article.
	Effectiveness and cost estimates were based on individual patient data.
	Results were presented in terms of the incremental or additional costs and outcomes of 1 intervention over another.
	Sensitivity analysis was presented if there was uncertainty.

*Methodologic criteria in column 2 were based on the cited references.

tional Library of Medicine associate fellows (Allmang NA, Koonce TY. Health services research topic searches. Bethesda [MD]: National Library of Medicine; 2000. Unpublished report). Three HSR experts had selected the journals in that list. The 3 journal lists were merged and duplicates deleted to yield the final list of 68 journals (for the complete merged list, see the online appendix at www.cmaj.ca/cgi/content/full/171/10/1179/DC1).

Manual review of the literature

Four research assistants reviewed each issue of the 68 journal titles for the calendar year 2000. Each journal article was read independently by 2 research assistants and coded for the following HSR categories, according to definitions derived using the MeSH scope notes (Table 1): appropriateness, process assessment, outcome assessment, CPGs, cost and economics. All original research and review articles that met the category definitions were evaluated for scientific merit on the basis of the criteria in Table 2, which were based on the "Users' guides to the medical literature" articles published in the *Journal of the American Medical Association*.[6-9] Although empirical evidence of design-related bias is not directly available for the HSR categories, research concerning diagnosis[13] and treatment[14] shows that studies with methodologic shortcomings may overestimate the accuracy or the effect being studied. To pass the criteria, an explicit statement relevant to each criterion had to appear in the article, and all criteria for the appropriate category had to be met. When disagreements arose between the assessments of the 2 research assistants, a third research assistant, blinded to the other assessments, reviewed the article in question. If the coding of the third appraiser agreed with the coding of 1 of the original reviewers, that coding was taken to be correct; otherwise, the article was referred to a more senior member of the research team, who reviewed all coding and determined the final classification.

Assessment of search terms

The candidate search terms were treated as "diagnostic tests" for sound studies, and the manual review of the literature was treated as the "gold standard." The concepts of diagnostic test evaluation and library science were used to determine the sensitivity, specificity and precision of MEDLINE searches as shown in Table 3. The sensitivity for a given topic was defined as the proportion of high-quality articles for that topic that were retrieved, specificity was the proportion of low-quality or nonrelevant articles that were not retrieved, and precision was the proportion of retrieved articles that were of high quality. Search performance was determined by an iterative computer program for each single term. Single terms that yielded sensitivity greater than 25% and specificity greater then 75% were used to form 2-term Boolean "or" strategy combinations. Two-term strategies that yielded sensitivity greater than 75% and specificity greater than 50% were used in 3-term Boolean "or" strategy development to optimize sensitivity. Two-term strategies that yielded sensitivity greater than 50% and specificity greater than 75% were used in 3-term Boolean "or" strategy development to optimize specificity. We did not test "and" combinations because of their predictably adverse effect on sensitivity. We also did not test "and not" combinations because, when we have tested this approach for clinical topics, the performance of the search strategies was not materially affected.

MEDLINE searches were conducted through Ovid (Ovid Technologies, New York; http://gateway2.ovid.com). For the defined sub-

Table 3: Formulas for calculating the sensitivity, specificity and precision of MEDLINE searches for detecting sound HSR studies*

MEDLINE search using search terms	Manual review	
	No. of studies meeting criteria	No. of studies not meeting criteria
No. of studies detected	a	b
No. of studies not detected	c	d
Total	a + c	b + d

*Sensitivity = a/(a + c), specificity = d/(b + d), precision = a/(a + b).

Table 4: Sample sizes for HSR categories

HSR category	No. of articles	No. (and %) meeting methodologic criteria
Appropriateness	20	5 (25)
Process assessment	251	48 (19)
Outcome assessment	266	209 (79)
Clinical practice guidelines	119	33 (28)
Cost	199	NA*
Economics	139	23 (17)
Total	994 (795†)	318 (40‡)

*No methodologic criteria were set for studies of the costs of health care services.
†Total excluding studies of cost.
‡Percentage based on total excluding studies of cost.

Table 5: Denominators for determining sensitivity, specificity and precision, as presented in Tables 6 to 8*

HSR category	No. of studies meeting criteria (a + c)	No. of studies not meeting criteria (b + d)
Appropriateness (A)		
All articles	20	25 916
Methodologically sound	5	25 931
Process assessment (PA)		
All articles	251	25 685
Methodologically sound	48	25 888
Outcome assessment (OA)		
All articles	266	25 670
Methodologically sound	209	25 727
Clinical practice guidelines (CPGs)		
All articles	119	25 817
Methodologically sound	33	25 903
A + PA + OA + CPGs†		
All articles	607	25 329
Methodologically sound	283	25 653
Cost (all articles)	199	25 737
Economics (methodologically sound articles)	23	25 913

*See Table 3 for related information.
†Totals are less than the sum of numbers for individual categories because some articles were classified in more than 1 category.

set of journal issues included in the database, we downloaded the full MEDLINE record, including full citation, abstract and MeSH terms. The MEDLINE records were then matched to the corresponding records in the hand-search file, by means of unique identifiers.

Results

The HSR database included 25 936 articles. Of these, 994 (3.8%) met our criteria for one or more of the HSR categories (Table 4). Of the 795 articles classified as relevant to appropriateness, process assessment, outcome assessment, CPGs and economics, 318 (40.0%) met our methodologic criteria (no methodologic criteria were applied for cost studies). The numbers of articles were adequate for precise estimates of search performance for all but the appropriateness category, which had only 20 articles in total and only 5 articles that met the criteria.

In total, 7445 unique search terms were tested, of which 5330 returned results. Predictably, single search terms had lower yields than 2-term strategies, but the difference between 2- and 3-term combinations was small. As expected, combining terms increased the sensitivity over single search terms. A somewhat unexpected finding was that some combinations of terms for each category of studies also led to increases in speci-

Table 6: Combinations of terms with the best sensitivity (and specificity ≥ 50%) for detecting HSR articles, for all articles and for methodologically sound studies in each HSR category*

HSR category	Exact Ovid search strategy†	Sensitivity (%)	Specificity (%)	Precision (%)
Appropriateness (A)				
All articles	appropriate:.tw. OR exp "organization and administration" OR exp delivery of health care	100	79.2	0.4
Methodologically sound	inappropriate.tw. OR exp records	100	97.4	0.7
Process assessment (PA)				
All articles	exp health services administration OR exp data collection OR hospital:.mp.	94.0	59.1	2.2
Methodologically sound	exp health services administration OR th.fs. OR "adjusted OR".tw.	97.9	66.7	0.5
	care.tw. OR th.fs. OR service:.tw.	93.8	78.8	0.8
Outcome assessment (OA)				
All articles	exp epidemiologic studies OR exp health services administration OR ep.xs.	94.0	53.0	2.0
Methodologically sound	exp health services administration OR exp epidemiologic studies OR ep.xs.	95.7	52.9	1.6
	exp epidemiologic studies OR ep.xs. OR exp "outcome and process assessment (health care)"	93.3	64.7	2.1
Clinical Practice Guidelines (CPGs)				
All classified	exp health services administration OR tu.xs. OR management.tw.	90.8	53.9	0.9
Methodologically sound	guideline:.tw. OR exp data collection OR recommend:.tw.	100	77.5	0.6
	guidelines.tw. OR practice guidelines.sh. OR recommend:.tw.	97.0	95.5	2.7
A + PA + OA + CPGs				
All articles	exp health services administration OR exp data collection OR exp epidemiologic studies	88.1	54.0	4.4
Methodologically sound	exp health services administration OR exp epidemiologic studies OR ep.xs.	95.4	53.1	2.2
	exp health services administration OR ep.xs. OR retrospective studies.sh.	92.6	60.1	2.5
Cost (all articles)‡	cost:.mp. OR risk:.mp. OR estimate:.tw.	96.0	77.6	3.2
	cost:.mp. OR ec.fs. OR exp health care costs	96.5	92.9	9.5
Economics (methodologically sound articles)§	costs.tw. OR cost effective.tw. OR economic.tw.	100	97.1	3.0

*See Table 5 for denominators used to calculate sensitivity, specificity and precision.
†Search strategies are in Ovid syntax and must be run exactly as presented. Colon = truncation, tw = text word (word or phrase appears in title or abstract), exp = exploded subject heading, mp = multiple posting (term appears in title, abstract or subject heading), th = therapy, fs = floating subheading, adjusted OR = adjusted odds ratio, ep = epidemiology, xs = exploded subheading, tu = therapeutic use, sh = subject heading, ec = economics.
‡No methodologic criteria applied.
§Data for all articles not available.

ficity and precision. Thus, for brevity, only 3-term strategies are presented here (see Table 5 for the denominators of the data presented in the tables for these strategies, as detailed below), unless 2-term or single-term strategies performed as well as the best 3-term strategy. The best sensitivities ranged from 95% to 100% for methodologically sound articles for all categories, including appropriateness studies (Table 6), but the estimate for the latter category was imprecise, as only 5 studies in the database met this criterion. Precision was 9.5% or less for all searches, a consequence of the low prevalence of HSR even in these selected journals and the suboptimal specificities for the most sensitive searches.

Terms that yielded the best specificity while maintaining sensitivity of 50% or more for each HSR category are presented in Table 7. In achieving the highest specificity for combined terms, sensitivity decreased in all HSR categories while precision rose somewhat.

The combinations of terms that optimized both sensitivity and specificity while minimizing the differences between the 2 measurements for each HSR category are presented in Table 8. These strategies provide the best separation of relevant from nonrelevant retrievals, but do so without regard for whether sensitivity or specificity is affected.

Interpretation

We have documented search strategies that can help to discriminate relevant from nonrelevant articles for a number of categories of importance to those interested in HSR. Those who are interested in all articles on a given topic and who have the time to sort out irrelevant articles will be best served by the most sensitive strategies (Table 6). Those with little time who are looking for "a few good articles" on a given topic will likely be best served by the most specific

Table 7: Combinations of terms with the best specificity (and sensitivity ≥ 50%) for detecting HSR articles, for all articles and for methodologically sound studies in each HSR category*

HSR category	Exact Ovid search strategy†	Sensitivity (%)	Specificity (%)	Precision (%)
Appropriateness (A)				
All articles	appropriateness.tw.	50.0	99.8	19.6
	appropriateness.tw. OR exp utilization review	65.0	99.6	11.1
Methodologically sound	appropriateness.tw.	60.0	99.8	5.9
	inappropriate.tw.	80.0	99.7	4.4
Process assessment (PA)				
All articles	service:.tw. OR hospital.tw.	54.6	90.6	5.5
Methodologically sound	practice:.tw. OR "adjusted OR".tw.	52.1	92.9	1.4
	exp patient care management OR exp insurance, health	66.7	91.8	1.5
Outcome assessment (OA)				
All articles	retrospective studies.sh. OR mortality.tw.	57.1	91.9	6.8
Methodologically sound	survival.tw. OR exp mortality	51.2	92.5	5.3
	retrospective studies.sh. OR mortality.tw.	63.2	91.8	5.9
Clinical Practice Guidelines (CPGs)				
All articles	guideline:.tw.	60.5	98.6	16.2
Methodologically sound	guideline adherence.sh. OR physician's practice patterns.sh.	51.5	98.8	5.2
	practice guidelines.tw. OR practice guidelines.sh.	69.7	98.7	6.3
A + PA + OA + CPGs				
All articles	hospital:.tw. OR retrospective studies.mp.	50.7	88.8	9.8
Methodologically sound	retrospective studies.sh. OR mortality.tw.	55.8	91.9	7.1
	retrospective studies.sh. OR mortality.tw. OR survival.tw.	65.0	88.8	6.0
Cost (all articles)‡	cost effectiveness:.tw. OR sav:.tw.	58.8	99.2	36.9
	cost effective:.tw. OR sav:.tw. OR cost-benefit analysis.sh.	71.9	98.3	24.3
Economics (methodologically sound articles)§	cost effective.tw. OR sensitivity analys:.tw.	56.5	99.3	7.1
	cost effective.tw. OR sensitivity analys:.tw. OR cost effectiveness.tw.	73.9	98.9	5.6

*See Table 5 for denominators used to calculate sensitivity, specificity and precision.
†Search strategies are in Ovid syntax and must be run exactly as presented.
‡No methodologic criteria applied.
§Data for all articles not available

strategies (Table 7). Use of these strategies is straightforward, as the National Library of Medicine has translated our most sensitive and most specific strategies for public use at www.nlm.nih.gov/nichsr/hedges/search.html. The best strategies for optimizing the trade-off between sensitivity and specificity are shown in Table 8. When sensitivity was maximized for combinations of terms, specificities rose considerably relative to those for individual search terms. For instance, for sensitive searches for high-quality appropriateness articles, the combination of terms resulted in marked increases, to near-perfect specificity. When specificity was maximized for combinations of terms, while keeping sensitivity of at least 50%, we observed very high specificities for almost all HSR categories. Search performance, including the trade-off between sensitivity and specificity, was generally comparable to that found for topics that are of more direct interest to clinicians, such as treatment, diagnosis, prognosis and etiology.[15] However, the methodologic standards for clinical topics are generally much higher, so that the literature retrieved provides more robust answers.

Few search filters have been developed to retrieve journal articles on a small range of topics of direct relevance to HSR. A pilot project created preliminary search strategies for economics and qualitative research in the HSR literature in 2000 (Allmang NA, Koonce TY. Health services research topic searches. Bethesda [MD]: National Library of Medicine; 2000. Unpublished report) but lacked a gold standard against which to assess the quality of the searches. Search filters developed for the National Health Service Economic Evaluation Database,[16] the Health Economic Evaluation Database[17] and the London School of Economics (LSE) Strategy,[18] which are designed to retrieve economic evaluation articles, were compared with one another, to generate a relative standard, giving estimates of sensitivity of 72% and specificity of 75% for the LSE strategy in MEDLINE.[18] Our findings for economics articles appear to be somewhat better but are not directly comparable, as our gold standard was a hand search. Additional filters have been designed to retrieve articles on outcome measurement[19] (just 3 strategies based on hand searches in

Table 8: Combinations of terms for optimizing sensitivity and specificity for detecting HSR articles, for all articles and for methodologically sound articles in each HSR category (based on abs[sensitivity – specificity] ≤ 1.5%)*

HSR category	Exact Ovid search strategy†	Sensitivity (%)	Specificity (%)	Precision (%)
Appropriateness (A)				
All articles	appropriate:.tw. OR exp utilization review OR exp delivery of health care	85.0	85.6	0.5
Methodologically sound	inappropriate:.mp. OR clinical:.tw. OR evaluation.tw.	80.0	81.0	0.1
Process assessment (PA)				
All articles	care.tw. OR hospital:.mp. OR service:.tw.	82.5	82.7	4.4
Methodologically sound	care.tw. OR "adjusted OR".tw. OR service:.tw.	85.4	86.0	1.1
Outcome assessment (OA)				
All articles	outcome.tw. OR retrospective stud:.mp. OR exp cohort studies	79.3	78.3	3.7
Methodologically sound	exp case-control studies OR survival:.mp. OR hospital:.tw.	82.8	82.2	3.6
Clinical Practice Guidelines (CPGs)				
All articles	guide:.tw. OR recommend:.tw. OR exp risk	84.9	84.2	2.4
Methodologically sound	exp "quality assurance (health care)" OR recommend:.tw. OR guideline adherence.sh.	93.9	94.1	2.0
A + PA + OA + CPGs				
All articles	care.tw. OR retrospective studies.sh. OR ep.xs.	72.9	72.4	6.0
Methodologically sound	hospital.mp. OR retrospective stud:.mp. OR th.fs.	78.1	78.4	3.8
Cost (all articles)‡	exp "costs and cost analysis" OR costs.tw. OR cost.tw.	95.0	95.6	14.3
Economics (methodologically sound articles)§	cost-benefit analysis.sh. OR costs.tw. OR cost effective.tw.	95.7	97.2	2.9

*See Table 5 for denominators used to calculate sensitivity, specificity and precision.
†Search strategies are in Ovid syntax and must be run exactly as presented.
‡No methodologic criteria applied.
§Data for all articles not available.

just 2 journals) and quality of care[20] (in which only precision was measured).

Our study had some limitations. First, we could not find secure methodologic features for the HSR categories of appropriateness and cost that lend themselves to retrieving the best studies. Second, the number of appropriateness articles in our database was small, giving rise to imprecise estimates of search performance for that category. Third, our database was not large enough to permit test–retest searches to validate the strategies. Fourth, we have not studied the effect of combining research filters with content terms (such as a disease, technology or type of health service) and thus cannot report on the characteristics of such searches; such a study would require considerably more resources than were available to us. Fifth, we tested only Ovid's search engine for MEDLINE; other search engines, including the PubMed search engine of the National Library of Medicine, may handle terms somewhat differently, with slightly differing results.

The best search strategies found in our research leave some room for improvement. Better search performance may require maturation of research methods for HSR, similar to those for some forms of clinical research, and better indexing. Improvements may also be possible through more sophisticated search strategies, for example, with more search terms, use of other Boolean operators ("and," "and not"), natural language processing and multivariate statistical techniques such as logistic regression and discriminant function analysis. In our limited experience with the use of other Boolean operators and logistic regression for clinical topics such as diagnostic tests,[21] we have observed trade-offs between sensitivity and specificity and no substantive improvements with more complex search strategies, but we have not attempted these approaches for HSR topics. We look forward to other researchers taking up the challenge of developing better search strategies for HSR.

This article has been peer reviewed.

From the Health Information Research Unit (Wilczynski, Haynes), the Department of Clinical Epidemiology and Biostatistics, (Haynes, Lavis), the Department of Medicine (Haynes), the Centre for Health Economics and Policy Analysis (Lavis) and the Department of Political Science (Lavis), McMaster University Health Sciences Centre, and the Department of Humanities and Social Sciences, Mohawk College (Ramkissoonsingh), Hamilton, Ont.; and the Department of Psychology, York University, Toronto, Ont. (Arnold-Oatley)

References

1. Ellrodt G, Cook DJ, Lee J, Hunt D, Weingarten S. Evidence-based disease management. *JAMA* 1997;278:1687-92.
2. Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ* 2002;324:1-7.
3. Lasker RD. Challenges to accessing useful information in health policy and public health: an introduction to a national forum held at the New York Academy of Medicine. *J Urban Health* 1998;75:779-84.
4. O'Carroll PW, Cahn MA, Auston I, Seden CR. Information needs in public health and health policy: results of recent studies. *J Urban Health* 1998;75:785-93.
5. Academy awarded new HSRProj contract. Acad Rep 2000;(2):8. Available: www.academyhealth.org/publications/academyhealthreports/oct00.pdf (accessed 2004 Sep 24).
6. Naylor CD, Guyatt GH. Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. Evidence-Based Medicine Working Group. *JAMA* 1996;275:1435-9.
7. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. Evidence-Based Medicine Working Group. *JAMA* 1996;275:554-8.
8. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? Evidence-Based Medicine Working Group. *JAMA* 1995;274:570-4.
9. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1997;277:1552-7.
10. Wilczynski NL, McKibbon KA, Haynes RB. Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature. *Medinfo* 2001;10(Pt 1):390-3.
11. Haynes RB. The origins and aspirations of ACP Journal Club [editorial]. *ACP J Club* 1991;114:A18.
12. Elixhauser A, Luce BR, Taylor WR, Reblando J. Health care CBA/CEA: an update on the growth and composition of the literature. *Med Care* 1993;31(7 Suppl):JS1-11.
13. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies in diagnostic tests. *JAMA* 1999;282:1061-6.
14. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
15. Wilczynski NL, Haynes Team. Robustness of empirical search strategies for clinical content in MEDLINE. *Proc AMIA Symp* 2002;:904-8.
16. NHS economic evaluation database (NHS EED) [on-line searchable database]. York (UK): University of York, National Health Service Centre for Reviews and Dissemination; updated 2004 Sep 24. Available: www.york.ac.uk/inst/crd/nhsdhp.htm (accessed 2004 Sep 24).
17. Health economic evaluation database [on-line searchable database]. London: Office of Health Economics; [date unknown]. Available: www.ohe.org (accessed 2004 Sep 7).
18. Sassi F, Archard L, McDaid D. Searching literature for health care economics evaluations: How systematic can we afford to be? *Med Care* 2002;40:387-94.
19. Brettle AJ, Long AF, Grant MJ, Greenhalgh J. Searching for information on outcomes: Do you need to be comprehensive? *Qual Health Care* 1998;7:163-7.
20. Balas EA, Stockham MG, Mitchell JA, Sievert ME, Ewigman BG, Boren SA. In search of controlled evidence for health care quality improvement. *J Med Syst* 1997;21:21-32.
21. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from MEDLINE: analytical survey. *BMJ* 2004;328(7447):1040.

Correspondence to: *Dr. R. Brian Haynes, Department of Clinical Epidemiology and Biostatistics, Room 2C10B, Health Sciences Centre, McMaster University Faculty of Health Sciences, 1200 Main St. W, Hamilton ON L8N 3J5; fax 905 577-0017; bhaynes@mcmaster.ca*

Members of the HSR Hedges team: *Nancy L. Wilczynski, R. Brian Haynes, John N. Lavis, Ann McKibbon, Douglas Morgan, Adrienne Stevens, Stephen Walter and Stephen Werre, McMaster University, Hamilton, Ont.; Ravi Ramkissoonsingh, Mohawk College, Hamilton, Ont.; Alexandra E. Arnold-Oatley, York University, Toronto, Ont.*