

## Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey

R Brian Haynes, K Ann McKibbin, Nancy L Wilczynski, Stephen D Walter, Stephen R Werre for the Hedges Team

### Abstract

**Objective** To develop and test optimal Medline search strategies for retrieving sound clinical studies on prevention or treatment of health disorders.

**Design** Analytical survey.

**Data sources** 161 clinical journals indexed in Medline for the year 2000.

**Main outcome measures** Sensitivity, specificity, precision, and accuracy of 4862 unique terms in 18 404 combinations.

**Results** Only 1587 (24.2%) of 6568 articles on treatment met criteria for testing clinical interventions. Combinations of search terms reached peak sensitivities of 99.3% (95% confidence interval 98.7% to 99.8%) at a specificity of 70.4% (69.8% to 70.9%). Compared with best single terms, best multiple terms increased sensitivity for sound studies by 4.1% (absolute increase), but with substantial loss of specificity (absolute difference 23.7%) when sensitivity was maximised. When terms were combined to maximise specificity, 97.4% (97.3% to 97.6%) was achieved, about the same as that achieved by the best single term (97.6%, 97.4% to 97.7%). The strategies newly reported in this paper outperformed other validated search strategies except for two strategies that had slightly higher specificity (98.1% and 97.6% *v* 97.4%) but lower sensitivity (42.0% and 92.8% *v* 93.1%).

**Conclusion** New empirical search strategies have been validated to optimise retrieval from Medline of articles reporting high quality clinical studies on prevention or treatment of health disorders.

### Introduction

Free worldwide internet access to the US National Library of Medicine's Medline service in early 1997 was followed by a 300-fold increase in searches (from 163 000 searches per month in January 1997 to 51.5 million searches per month in December 2004<sup>1</sup>), with direct use by clinicians, students, and the general public growing faster than use mediated by librarians.

If large electronic bibliographic databases such as Medline are to be helpful to clinical users, clinicians must be able to retrieve articles that are scientifically sound and directly relevant to the health problem they are trying to solve, without missing key studies or retrieving excessive numbers of preliminary, irrelevant, outdated, or misleading reports. Few clinicians, however, are trained in search techniques. One approach to enhance the effectiveness of searches by clinical users is to develop search filters ("hedges") to improve the retrieval of clinically relevant and scientifically sound reports of studies from Medline and similar bibliographic databases.<sup>2-7</sup> Hedges can be created with appropri-

ate disease content terms combined ("ANDed") with medical subject headings (MeSH), explosions (px), publication types (pt), subheadings (sh), and textwords (tw) that detect research design features indicating methodological rigour for applied healthcare research. For instance, combining clinical trial (pt) AND myocardial infarction in PubMed brings the retrieval for myocardial infarction down by a factor of 13 (from 116 199 to 8956 articles) and effectively removes case reports, laboratory and animal studies, and other less rigorous and extraneous reports.

In the early 1990s, our group developed Medline search filters for studies of the cause, course, diagnosis, or treatment of health problems, based on a small subset of 10 clinical journals.<sup>8</sup> These strategies were adapted for use in the Clinical Queries feature in PubMed and other services. In this paper we report improved hedges for retrieving studies on prevention and treatment, developed on a larger number of journals (n=161) in a more current era (2000) than previously reported.<sup>9</sup>

### Methods

Our methods are detailed elsewhere.<sup>10 11</sup> Briefly, research staff hand searched each issue of 161 clinical journals indexed in Medline for the year 2000 to find studies on treatment that met the following criteria: random allocation of participants to comparison groups, outcome assessment for at least 80% of those entering the investigation accounted for in one major analysis for at least one follow-up assessment, and analysis consistent with study design. Search strategies were then created and tested for their ability to retrieve articles in Medline that met these criteria while excluding articles that did not.

Table 1 shows the sensitivity, specificity, precision, and accuracy of single term and multiple term Medline search strategies that we determined. The sensitivity for a given strategy is defined as the proportion of articles retrieved that are scientifically sound and clinically relevant (high quality articles); specificity is the proportion of lower quality articles (did not meet criteria) that are not retrieved; precision is the proportion of retrieved articles that meet criteria (equivalent to positive predictive value in diagnostic test terminology); and accuracy is the proportion of all articles that are correctly dealt with by the strategy (articles that met criteria and were retrieved plus articles that did not meet criteria and were not retrieved divided by all articles in the database).

After extensive attempts, a small fraction (n=968, 2%) of citations downloaded from Medline could not be matched to the handsearched data. As a conservative approach, unmatched citations that were detected by a given search strategy were included in cell b of the analysis in table 1 (leading to slight underestimates of the precision, specificity, and accuracy of the

**Table 1** Formula for calculating sensitivity, specificity, precision, and accuracy of Medline searches for detecting sound clinical studies

Search terms	Manual review	
	Meets criteria	Does not meet criteria
Detected	a	b
Not detected	c	d
	a+c	b+d

Sensitivity= $a/(a+c)$ ; precision= $a/(a+b)$ ; specificity= $d/(b+d)$ ; accuracy= $(a+d)/(a+b+c+d)$ . All articles classified during manual review of literature= $(a+b+c+d)$ .

search strategy). Similarly, unmatched citations that were not detected by a search strategy were included in cell d of the table (leading to slight overestimates of the specificity and accuracy of the strategy).

### Manual review of the literature

The figure illustrates the steps involved in the data collection and analysis stages. Six research assistants completed rigorous calibration exercises for application of methodological criteria to articles to determine if the article was methodologically sound. Inter-rater agreement for the classification of articles, corrected for chance agreement, exceeded 80%.<sup>12</sup>

### Collecting search terms

To construct a comprehensive set of possible search terms, we listed MeSH terms and textwords related to study criteria and then sought input from clinicians and librarians through interviews and requests at meetings and conferences and through electronic mail, review of published and unpublished search strategies from other groups, and requests to the National Library of Medicine. We compiled a list of 4862 unique terms (data not shown). All terms were tested using the Ovid Technologies searching system. Search strategies developed using Ovid were subsequently translated by the National Library of Medicine for use in the Clinical Queries interface of PubMed and reviewed by RBH.

### Data collection

Manual ratings of articles were recorded on data collection forms along with bibliographic information and database specific unique identifiers. Each journal title was searched in Medline for 2000, and the full Medline records (including citation, abstract, MeSH terms, and publication types) were cap-

ured for all articles. Medline data were then linked with the manual review data.

### Testing strategies

We randomly divided treatment and prevention articles that met criteria in the manual review database into development and validation datasets (60% and 40%). Sensitivity, specificity, precision, and accuracy were calculated for each term in the development subset and then validated in the rest of the database. For a given purpose category, we incorporated individual search terms with sensitivity greater than 25% and specificity greater than 75% into the development of search strategies that included a combination of two or more terms. All combinations of terms used the boolean OR—for example, “random OR controlled”. (The boolean AND was not used because this strategy invariably compromised sensitivity.)

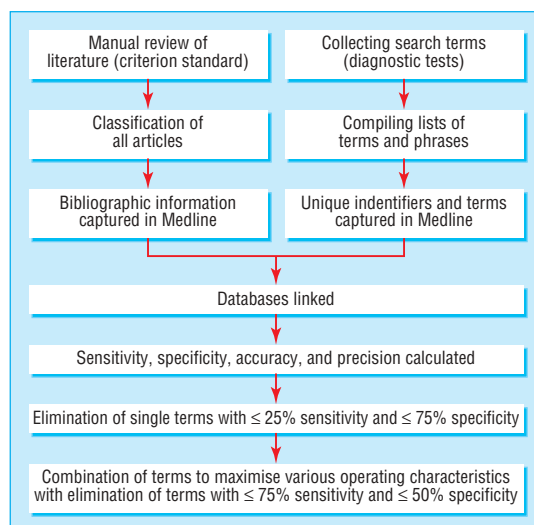
For the development of multiple term search strategies to optimise either sensitivity or specificity, we tested all two term search strategies with sensitivity at least 75% and specificity at least 50%. For optimising accuracy, two term search strategies with accuracy greater than 75% were considered for multiple term development. Overall, we tested 18 404 multiple term search strategies. Search strategies were also developed that optimised combined sensitivity and specificity (by keeping the absolute difference between sensitivity and specificity less than 1%, if possible).

To attempt to increase specificity without compromising sensitivity, we used terms with low sensitivity but appreciable specificity to NOT out citations (for example, randomised controlled trial.pt OR randomized.mp OR placebo.mp NOT retrospective studies.mp. (where pt=publication type; mp= multiple posting—term appears in title, abstract, or MeSH heading)). We also used logistic regression analysis models that included terms in a stepwise manner and also NOTed out terms with a regression coefficient less than  $-2.0$ .

We compared strategies that maximised each of sensitivity, specificity, precision, and accuracy for both development and validation datasets with 19 previously published strategies. We chose strategies that had been tested against an ideal method such as a hand search of the published literature and for which most Medline records were from 1990 forward, to reflect major changes in the classification of clinical trials by the National Library of Medicine. These changes included new MeSH definitions (for example, “cohort studies” was introduced in 1989 and “single-blind method” in 1990) and publication types (for example, “clinical trial (pt)” and “randomized controlled trial (pt)”, which were instituted in 1991). Six papers<sup>2-7</sup> and one library website<sup>13</sup> provided a total of 19 strategies to test, including the strategy advocated by the Cochrane Collaboration in their handbook ([www.cochrane.dk/cochrane/handbook/hbookAPPENDIX\\_5C\\_OPTIMAL\\_SEARCH\\_STRAT.htm](http://www.cochrane.dk/cochrane/handbook/hbookAPPENDIX_5C_OPTIMAL_SEARCH_STRAT.htm)).<sup>2</sup>

### Results

We included 49 028 articles in the analysis; 6568 articles (13.4%) were classified as original studies evaluating a treatment, of which 1587 (24.2%) met our methodological criteria. Overall, 3807 of 4862 proposed unique terms retrieved citations from Medline that could be used in assessment of terms. The development and validation datasets for assessing retrieval strategies included articles that passed and did not pass treatment criteria (930 and 29 397 articles, respectively, for the development dataset; 657 and 19 631 articles for the validation dataset). The validation dataset provided differences in performance that were



Steps in data collection to determine optimal retrieval of articles on treatment from Medline

**Table 2** Best single terms for high sensitivity searches, high specificity searches, and searches that optimise balance between sensitivity and specificity for retrieving studies of treatment

Search strategy in Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
High sensitivity*				
Clinical trial.mp.pt.	95.2 (93.8 to 96.5)	94.1 (93.8 to 94.4)	34.5 (32.7 to 36.4)	94.1 (93.9 to 94.4)
High specificity†				
Randomized controlled trial.pt.	92.8 (91.1 to 94.5)	97.6 (97.4 to 97.7)	55.5 (52.8 to 57.8)	97.4 (97.2 to 97.6)
Optimising sensitivity and specificity‡				
Randomized controlled trial.pt.	92.8 (91.1 to 94.5)	97.6 (97.4 to 97.7)	55.5 (52.8 to 57.8)	97.4 (97.2 to 97.6)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type.

\*Keeping specificity >50%.

†Keeping sensitivity >50%.

‡abs(sensitivity-specificity)<1%.

statistically significant in only three of 36 comparisons, the greatest of which was 1.1% for one set of specificities (data not shown).

Table 2 shows the operating characteristics for the single terms with the highest sensitivity and the highest specificity. The accuracy is driven by the specificity and thus the term with the best accuracy when keeping sensitivity more than 50% was “randomized controlled trial.pt.”. The single term that yielded the best precision while keeping sensitivity more than 50% was also “randomized controlled trial.pt.”, and this strategy also gave the optimal balance of sensitivity and specificity.

For strategies combining up to three terms, those yielding the highest sensitivity, specificity, and accuracy are shown in tables 3 to 5. Some two term strategies outperformed one term and multiple term strategies (table 5). Table 6 shows the top three search strategies optimising the trade-off between sensitivity and specificity.

Table 7 shows the best combination of terms for optimising the trade-off between sensitivity and specificity when using the boolean NOT to eliminate terms with the lowest sensitivity. Non-significant differences were shown when citations retrieved by

the three terms “review tutorial.pt.”, “review academic.pt.”, and “selection criteri.:tw.” were removed from the strategy that optimised sensitivity and specificity.

After the two term and three term computations, search strategies with sensitivity more than 50% and specificity more than 95% were further evaluated by adding search terms selected using logistic regression modelling. Initially, candidate terms for addition to the base strategy were ordered with the most significant first, using stepwise logistic regression, and then added to the model sequentially. The resulting logistic function (data not shown) determined the association between the predicted probabilities and observed responses. We selected the best one term, two term, three term, and four term strategies. Two were already evaluated (“randomized controlled trial.mp.” OR “randomized controlled trial.pt.” in table 4 and “randomized controlled trial.mp.” OR “randomized controlled trial.pt.” OR “double-blind.tw.” in table 5). The other two strategies are listed in table 8: both had high performance. We next took the 13 terms that had regression coefficients less than  $-2.0$  (“predict.tw.”, “predict.mp.”, “economic.tw.”, “economic.mp.”, “survey.tw.”, “survey.mp.”, “hospi-

**Table 3** Top three search strategies yielding highest sensitivity (keeping specificity >50%) with combinations of terms

Search strategy in Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
Clinical trial.mp. OR clinical trial.pt. OR random.:mp. OR tu.xs.	99.3 (98.7 to 99.8)	70.4 (69.8 to 70.9)	9.9 (9.3 to 10.5)	71.3 (70.8 to 71.8)
Clinical trial.pt. OR random.:tw. OR tu.xs.	99.1 (98.6 to 99.7)	71.0 (70.4 to 71.5)	10.0 (9.4 to 10.6)	71.8 (71.3 to 72.4)
Clinical trial.pt. OR random.:mp. OR tu.fs.	98.9 (98.3 to 99.6)	79.7 (79.3 to 80.2)	13.8 (12.9 to 14.6)	80.3 (79.9 to 80.8)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type; tu=therapeutic use subheading; xs=explode subheading; :-truncation; tw=textword (word or phrase appears in title or abstract); fs=floating subheading.

**Table 4** Top three search strategies yielding highest specificity (keeping sensitivity >50%) based on combinations of up to three terms

Search strategy in Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
Randomized controlled trial.mp. OR randomized controlled trial.pt.	93.1 (91.5 to 94.8)	97.4 (97.3 to 97.6)	54.4 (52.0 to 56.8)	97.3 (97.1 to 97.5)
Randomized controlled trial.pt. OR Double-blind method.sh.	93.7 (92.1 to 95.2)	97.3 (97.1 to 97.5)	52.9 (50.4 to 55.3)	97.2 (97.0 to 97.4)
Randomized controlled trial.pt. OR Double-blind.mp.	93.8 (92.2 to 95.3)	97.2 (97.0 to 97.4)	52.5 (50.1 to 54.9)	97.1 (96.9 to 97.3)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type; sh=MeSH.

**Table 5** Top three search strategies yielding highest accuracy (keeping sensitivity >50%) based on combinations of up to three terms

Search strategy using Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
Randomized controlled trial.mp. OR randomized controlled trial.pt.	93.1 (91.5 to 94.8)	97.4 (97.3 to 97.6)	54.4 (52.0 to 56.8)	97.3 (97.1 to 97.5)
Randomized controlled trial.pt. OR double-blind method.sh.	93.7 (92.1 to 95.2)	97.3 (97.1 to 97.5)	52.9 (50.4 to 55.3)	97.2 (97.0 to 97.4)
Randomized controlled trial.pt. OR double-blind.tw.	93.7 (92.2 to 95.3)	97.2 (97.0 to 97.4)	52.5 (50.1 to 54.9)	97.1 (96.9 to 97.3)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type; sh=MeSH; tw=textword (word or phrase appears in title or abstract).

**Table 6** Top three search strategies for optimising sensitivity and specificity (based on absolute difference (sensitivity—specificity) <1%)

Search strategy using Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
Randomized controlled trial.pt. OR randomized.mp. OR placebo.mp.	95.8 (94.5 to 97.1)	95.0 (94.8 to 95.3)	38.5 (36.5 to 40.5)	95.0 (94.8 to 95.3)
Randomized controlled trial.mp. OR randomized controlled trial.pt. OR randomized.tw. OR placebo:.tw.	95.8 (94.5 to 97.1)	95.0 (94.7 to 95.2)	38.4 (36.4 to 40.4)	95.0 (94.8 to 95.3)
Randomized controlled trial.pt. OR randomized.tw. OR placebo:.tw.	95.8 (94.5 to 97.1)	95.0 (94.7 to 95.2)	38.4 (36.4 to 40.4)	95.0 (94.8 to 95.3)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type; :=truncation; tw=textword (word or phrase appears in title or abstract).

tal mortality.mp,tw”, “hospital mortalit.mp.”, “accuracy:.tw.”, “accuracy.tw.”, “accuracy.mp.”, “explode bias (epidemiology)”, and “longitudinal.tw.”) and NOTed these terms out of the four term search strategy to determine if these terms would improve the operating characteristic values (table 8, last row). We found a small but insignificant decrease in sensitivity and increases in specificity, precision, and accuracy.

We compared our best strategies for maximising sensitivity (sensitivity >99% and specificity >70%) and for maximising specificity while maintaining a high sensitivity (sensitivity >94% and specificity >97%). To ascertain if the less sensitive strategy (which had a much greater specificity) would miss important articles, we assessed the methodologically sound articles that had not been retrieved by the less sensitive strategy, using studies from the four major medical journals (*BMJ*, *JAMA*, *Lancet*, and *New England Journal of Medicine*). In total, 32 articles were missed by the less sensitive search, of which four were from these four

journals. A practising clinician with training in methods for health research found only one of the four articles to be of substantial clinical importance.<sup>14</sup> The indexing terms for this randomised controlled trial did not include “randomized controlled trial(pt)”. When we contacted the National Library of Medicine about indexing for this article, the article was reindexed and now the “missing” article would be retrieved.

We used our data to test 19 published strategies<sup>2-7 13</sup> and we compared these with the best strategies for optimising sensitivity and specificity. The published strategies had a sensitivity range of 1.3% to 98.8% on the basis of our handsearched data. All of these were lower than our best sensitivity of 99.3%. The specificities for the published strategies ranged from 63.3% to 96.6%. Two strategies from Dumbrique<sup>6</sup> outperformed our most specific strategy (specificity of 98.1% and 97.6% versus our 97.4%). Both of these strategies had a lower sensitivity than did our search strategy with the best specificity (42.0% and 92.8% v 93.1%).

**Table 7** Best combination of terms for optimising the trade-off between sensitivity and specificity in Medline when adding the boolean AND NOT

Search strategy using Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
Randomized controlled trial.pt. OR randomized.mp. OR placebo.mp. AND NOT (review tutorial.pt. OR search.tw. OR review academic.pt. OR selection criteri:.tw.)	95.8 (94.5 to 97.1)	96.1 (95.8 to 96.3)	44.4 (42.1 to 46.4)	96.0 (95.8 to 96.3)

pt=publication type; mp=multiple posting (term appears in title, abstract, or MeSH heading); :=truncation; tw=textword (word or phrase appears in title or abstract); :=truncation.

**Table 8** Top three term and four term search strategies using logistic regression techniques

Search strategy in Ovid format	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	Accuracy (%) (95% CI)
randomized controlled trial.mp. OR randomized controlled trial.pt. OR double-blind:.tw. OR random: assigned.mp.	95.2 (93.8 to 96.5)	96.9 (96.7 to 97.1)	50.3 (48.0 to 52.7)	96.9 (96.7 to 97.1)
randomized controlled trial.mp. OR randomized controlled trial.pt. OR double-blind:.tw. OR random: assigned.mp. OR randomized trial.mp.	95.5 (94.2 to 96.8)	96.8 (96.6 to 97.0)	49.4 (47.1 to 51.8)	96.8 (96.6 to 97.0)
randomized controlled trial.mp. OR randomized controlled trial.pt. OR double-blind:.tw. OR random: assigned.mp. OR randomized trial.mp. NOT all terms with coefficients < -2.0.	92.8 (91.1 to 94.5)	97.1 (97.0 to 97.3)	51.5 (49.1 to 53.9)	97.0 (96.8 to 97.2)

mp=multiple posting (term appears in title, abstract, or MeSH heading); pt=publication type; :=truncation; tw=textword (word or phrase appears in title or abstract).

**Table 9** Comparison of strategies from 1991 with newly developed strategies from 2000, compiled using 2000 data

Year	Approach	Strategy in Ovid format	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
1991	Maximise sensitivity	Randomized controlled trial.pt. OR drug therapy.fs. OR therapeutic use.fs. OR random:.tw.	98.3	79.4	13.4	78.6
2000	Maximise sensitivity	Clinical trial.pt. OR random:.mp. OR therapeutic use.fs.	98.9	79.7	13.8	80.3
1991	Maximise specificity	Double.tw. AND blind.tw. OR placebo:.tw.	42.3	98.1	42.2	96.3
2000	Maximise specificity	Randomized controlled trial.mp. OR randomized controlled trial.pt.	93.1	97.4	54.4	97.3

pt=publication type; fs=floating subheading; :=truncation; tw=textword (word or phrase appears in title or abstract); mp=multiple posting (term appears in title, abstract, or MeSH heading).

## Discussion

We have presented a variety of tested search strategies for retrieval of high quality and clinically ready studies of treatments. This research updates our previous hedges published in 1991,<sup>8</sup> calibrated using 10 internal and general medicine journals. When these 1991 strategies were tested in the 2000 database, the performance of the 2000 strategies was slightly better for the strategy that maximised sensitivity and considerably better for the strategy that maximised specificity (table 9). For example, when sensitivity was maximised, the sensitivity, specificity, precision, and accuracy were all slightly higher than for the most sensitive strategy in 1991. When specificity was optimised, the 2000 strategy had much higher sensitivity than the 1991 strategy (93.1% *v* 42.3%, respectively), with a small loss in specificity and gains in precision and accuracy.

No one search strategy will perform perfectly, for several reasons. Indexing inconsistencies affect retrievals, as shown by the need to reindex in the study by Julien et al.<sup>14</sup> Indexing terms and methods are modified over time and few changes are implemented retrospectively. Indexers also choose only a small number of terms for each item they index, and many of these terms have similar meanings—for example, “randomized controlled trials” and “clinical trials” as MeSH and “randomized controlled trial” and “clinical trial” as publication types. Methods and their naming also change over time, and authors may also be imprecise in their description of methods and results, affecting retrievals that are based on textwords in the titles and abstracts. The model we used for testing search strategies defines the constant features of these strategies (their sensitivity and specificity), and these strategies can be expected to perform the same way in the entire Medline database, as shown by their performance in the validation database in our study and by the robustness of the 1991 strategies when retested in our much larger 2000 database.<sup>9</sup> The precision of searches, however, depends on the concentration of relevant articles in the database. We selected clinical journals to calibrate the search strategies, but Medline contains many non-clinical journals. Thus, the concentration of high quality treatment studies will be less in the full Medline database, and the precision of searches will be less accordingly. This is a problem that warrants further attention.

Searchers who want retrieval with little non-relevant material—for instance, practising clinicians with little time to sort through many irrelevant articles can choose strategies with high specificity. For those interested in comprehensive retrievals, for instance researchers conducting systematic reviews or those searching for clinical topics with few citations, strategies with higher sensitivity will be more appropriate. Regardless of the strategy used, the most effective way to harness these strategies is to have them embedded within searching systems. The most sensitive and most specific search strategies reported here have been implemented in the Clinical Queries search screen ([www.ncbi.nlm.nih.gov/80/entrez/query/static/clinical.html](http://www.ncbi.nlm.nih.gov/80/entrez/query/static/clinical.html)) and by Ovid Technologies ([www.ovid.com](http://www.ovid.com)), and the optimal strategy has been added to Skolar ([www.skolar.com](http://www.skolar.com)).

The Hedges Team includes Angela Eady, Brian Haynes, Susan Marks, Ann McKibbin, Doug Morgan, Cindy Walker-Dilks, Stephen Walter, Stephen Werre, Nancy Wilczynski, and Sharon Wong, all at McMaster University Faculty of Health Sciences.

Contributors: RBH and NLW prepared grant submissions for this project. All authors drafted, commented on, and approved the final manuscript, and supplied intellectual content to the collection and analysis of the data. KAM and NLW participated in the data collection and, with the addition of RBH, were involved in data analysis and staff supervision. RBH is guarantor for the paper.

### What is already known on this topic

Many clinicians and researchers conduct Medline searches independently but lack skills to do this well

A barrier to searching for evidence in Medline is the difficulty selecting an optimal strategy to search for information

### What this study adds

Special Medline search strategies were developed and tested that retrieve 99% of scientifically sound therapy articles

Clinicians can use the most specific search when looking for a few sound articles on a topic

Researchers can use the most sensitive search when carrying out a comprehensive search for trials for their systematic reviews

These strategies have been automated for use in PubMed (Clinical Queries screen), and Ovid Technology's Medline and Skolar services

Funding: National Institutes of Health (grant RO1 LM06866-01).

Competing interests: None declared.

Ethical approval: Not required.

- 1 National Library of Medicine, Bibliographic Services Division. Number of Medline searches. [www.nlm.nih.gov/bsd/medline\\_growth\\_508.html](http://www.nlm.nih.gov/bsd/medline_growth_508.html) (accessed 3 Jul 2003).
- 2 Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol* 2002;31:150-3.
- 3 Nwosu CR, Khan KS, Chien PF. A two-term Medline search strategy for identifying randomized trials in obstetrics and gynecology. *Obstet Gynecol* 1998;91:618-22.
- 4 Marson AG, Chadwick DW. How easy are randomized controlled trials in epilepsy to find on Medline? The sensitivity and precision of two Medline searches. *Epilepsia* 1996;37:377-80.
- 5 Adams CE, Power A, Frederick K, LeFebvre C. An investigation of the adequacy of Medline searches for randomized controlled trials (RCTs) of the effects of mental health care. *Psychol Med* 1994;24:741-8.
- 6 Dumbrigue HB, Esquivel JF, Jones JS. Assessment of Medline search strategies for randomized controlled trials in prosthodontics. *J Prosthodont* 2000;9:8-13.
- 7 Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993;No 33.
- 8 Haynes RB, Wilczynski N, McKibbin KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinical sound studies in Medline. *J Am Med Assoc* 1994;1:447-58.
- 9 Wilczynski NL, Haynes RB. Robustness of empirical search strategies for clinical content in Medline. *Proc AMIA Symp* 2002;904-8.
- 10 Haynes RB, Wilczynski NC for the Hedges Team. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline. *BMJ* 2004;328:1040-2.
- 11 Montori VM, Wilczynski NL, Morgan D, Haynes RB for the Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: an analytical survey. *BMJ* 2005;330:68-73.
- 12 Wilczynski NL, McKibbin 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.
- 13 University of Rochester Medical Center. Edward G. Miner Library. Evidence-based filters for Ovid Medline. [www.urmc.rochester.edu/Miner/links/eBMLinks.html#TOOLS](http://www.urmc.rochester.edu/Miner/links/eBMLinks.html#TOOLS) (accessed 22 May 2003).
- 14 Julien JP, Bijker N, Fentiman IS, Peterse JL, Delledonne V, Rouanet P, et al. Radiotherapy in breast-conserving treatment for ductal carcinoma in situ: first results of the EORTC randomised phase III trial 10853. EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. *Lancet* 2000;355:528-33. (Accepted 31 March 2005)

doi 10.1136/bmj.38446.498542.8F

Health Information Research Unit, McMaster University, Hamilton, ON, Canada L8N 3Z5

R Brian Haynes *chief*  
Stephen R Werre *research associate*

School of Graduate Studies, McMaster University  
Nancy L Wilczynski *doctoral candidate*

Department of Clinical Epidemiology and Biostatistics, McMaster University  
Stephen D Walter *professor*

Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA  
K Ann McKibbon *doctoral candidate*

Correspondence to: R B Haynes [bhaynes@mcmaster.ca](mailto:bhaynes@mcmaster.ca)