

Research article

Open Access

## Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey

Nancy L Wilczynski<sup>1</sup>, R Brian Haynes\*<sup>1,2</sup> and the Hedges Team

Address: <sup>1</sup>Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada and <sup>2</sup>Department of Medicine, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

Email: Nancy L Wilczynski - wilczyn@mcmaster.ca; R Brian Haynes\* - bhaynes@mcmaster.ca; the Hedges Team -

\* Corresponding author

Published: 09 June 2004

Received: 03 February 2004

*BMC Medicine* 2004, **2**:23

Accepted: 09 June 2004

This article is available from: <http://www.biomedcentral.com/1741-7015/2/23>

© 2004 Wilczynski et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Clinical end users of MEDLINE have a difficult time retrieving articles that are both scientifically sound and directly relevant to clinical practice. Search filters have been developed to assist end users in increasing the success of their searches. Many filters have been developed for the literature on therapy and reviews but little has been done in the area of prognosis. The objective of this study is to determine how well various methodologic textwords, Medical Subject Headings, and their Boolean combinations retrieve methodologically sound literature on the prognosis of health disorders in MEDLINE.

**Methods:** An analytic survey was conducted, comparing hand searches of journals with retrievals from MEDLINE for candidate search terms and combinations. Six research assistants read all issues of 161 journals for the publishing year 2000. All articles were rated using purpose and quality indicators and categorized into clinically relevant original studies, review articles, general papers, or case reports. The original and review articles were then categorized as 'pass' or 'fail' for methodologic rigor in the areas of prognosis and other clinical topics. Candidate search strategies were developed for prognosis and run in MEDLINE – the retrievals being compared with the hand search data. The sensitivity, specificity, precision, and accuracy of the search strategies were calculated.

**Results:** 12% of studies classified as prognosis met basic criteria for scientific merit for testing clinical applications. Combinations of terms reached peak sensitivities of 90%. Compared with the best single term, multiple terms increased sensitivity for sound studies by 25.2% (absolute increase), and increased specificity, but by a much smaller amount (1.1%) when sensitivity was maximized. Combining terms to optimize both sensitivity and specificity achieved sensitivities and specificities of approximately 83% for each.

**Conclusion:** Empirically derived search strategies combining indexing terms and textwords can achieve high sensitivity and specificity for retrieving sound prognostic studies from MEDLINE.

### Background

Searching for the best evidence in MEDLINE can be difficult as it involves searching through over 5,000 journals

with an estimated 8,000 citations entered on a weekly basis. The task is increasingly difficult because advances in health care practice are published in a wide array of

journals, mixed with many preliminary studies. This explosion and scattering of information makes it difficult for clinicians to keep up to date with advances in health care [1,2] resulting in most researchable information needs being unmet [3]. Clinicians are expected to use the most relevant evidence from research but to do so they must be able to identify the best evidence reliably and efficiently. Even clinicians who support evidence-based medicine in principle often believe they do not do this in practice [4]. When they do try to find research evidence, practitioners do not search the medical literature very effectively [5]. One of the six most salient obstacles identified by doctors when attempting to answer questions about patient care is difficulty in selecting an optimal strategy to search for information [6]. If databases such as MEDLINE are to be helpful to clinicians, they 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 irrelevant or misleading studies.

One method of helping clinical searchers is to develop methodologic search filters to improve the retrieval of clinically relevant and scientifically sound study reports from databases such as MEDLINE. In MEDLINE, filters are created by adding, to disease content terms, Medical Subject Headings (MeSH), explosions (exp), publication types (pt), subheadings (xs or fs), and textwords (tw) that detect research design features indicating methodologic rigor for applied health care research; for example, 'myocardial infarction and (randomized controlled trial (pt) or clinical trial (pt))'. The use of these types of methodologic search filters has been advocated [7], and filters have been developed to improve the accuracy of searching for such studies [8-10]. Most of the studies have focused on information retrieval for therapy and diagnostic articles as well as systematic reviews. Little work has been done in the area of prognosis and to our knowledge, our previous study [11,12] was the only one in which search strategies for prognosis were empirically tested.

In the early 1990s, our group developed search filters on a subset of 10 journals for four types of journal articles: therapy, diagnosis, prognosis and causation [11,12]. These strategies have been adapted for use in the Clinical Queries interface of MEDLINE [13]. We are updating this research in the publishing year 2000 and have expanded the list of journals to 161. The robustness of the search strategies developed in 1991 for detecting clinical content in MEDLINE in the year 2000 has already been reported [14]. In this paper, we report on the information retrieval properties of a broader range of single terms and combinations of terms in MEDLINE for identifying methodologically sound studies on the prognosis of health

disorders, developed on a much larger set of journals than previously.

## Methods

The study compared the retrieval performance of methodologic search terms and phrases in MEDLINE with a manual review of each article for each issue of 161 journal titles for the year 2000. MeSH terms and textwords related to research design features were run as search strategies. The search strategies were treated as 'diagnostic tests' for sound studies and the manual review of the literature was treated as the 'gold standard'. The sensitivity, specificity, precision, and accuracy of MEDLINE searches were determined. Sensitivity for a given topic is defined as the proportion of high quality articles for that topic that are retrieved; specificity is the proportion of low quality articles not retrieved; precision is the proportion of retrieved articles that are of high quality; and accuracy is the proportion of all articles that are correctly classified.

Six research assistants hand searched the 161 journals titles for the year 2000, and applied methodologic criteria to each item in each issue to determine if the article was methodologically sound for seven purpose categories (two other types of articles, cost and qualitative studies, were also classified but had no rigor criteria). All purpose category definitions and corresponding methodologic rigor were outlined in a previous paper [15]. The focus of the strategies is to help clinicians retrieve methodologically sound study reports, as patient care decisions should be based on good quality evidence. The methodologic criteria applied for studies of prognosis were as follows: inception cohort of individuals all initially free of the outcome of interest; follow-up of at least 80% of patients until the occurrence of a major study end point or to the end of the study; and analysis consistent with study design.

The selection of the 161 journal titles reviewed was based on recommendations of clinicians and librarians, Science Citation Index Impact Factors provided by the Institute for Scientific Information, and ongoing assessment of their yield of studies and reviews of scientific merit and clinical relevance for the disciplines of internal medicine, general medical practice, mental health, and general nursing practice (list of journals provided by the authors upon request). Examples of the 161 journal titles included in the hand search are *Addiction*, *Age & Ageing*, *BMJ*, *JAMA*, *Lancet*, *New Journal of Medicine*, *Pediatrics*, *Public Health Nursing*, and *Stroke*. Research staff were rigorously calibrated prior to reviewing the 2000 literature and inter-rater agreement for application of all criteria exceeded 80% beyond chance [15].

**Table 1: Single term with the best sensitivity (keeping specificity  $\geq 50\%$ ), best specificity (keeping sensitivity  $\geq 50\%$ ), and best optimization of sensitivity and specificity (based on the lowest possible absolute difference between sensitivity and specificity) for detecting studies of prognosis in MEDLINE in 2000**

Search term OVID search*	Sensitivity (%) Development Validation Diff (95% CI)†	Specificity (%) Development Validation Diff (95% CI)†	Precision (%) Development Validation Diff (95% CI)†	Accuracy (%) Development Validation Diff (95% CI)†
‡exp epidemiologic studies	64.9 73.4 8.5 (-4.6 to 21.7)	78.6 79.1 0.5 (-0.3 to 1.2)	1.1 1.4 0.3 (-0.2 to 0.7)	78.6 79.1 0.5 (-0.3 to 1.2)

\*The search strategy is reported using Ovid's search engine syntax for MEDLINE. The PubMed syntax is epidemiologic studies [MeSH]. †Diff = Difference, comparing the development and validation data sets using the iterative method of Miettinen and Nurminen for two independent binomial proportions. None of the differences were statistically significant. ‡exp = explode, a search term that automatically includes closely related indexing terms.

An initial list of MeSH terms and textwords was compiled. Input was then sought from clinicians and librarians in the United States and Canada through interviews of known searchers, requests at meetings and conferences, and requests to the National Library of Medicine. Individuals were asked to identify which terms or phrases they used when searching for studies of prognosis, causation, diagnosis, treatment, economics, clinical prediction guides, reviews, costs, and of a qualitative nature. Terms could be from MeSH, including publication types and subheadings, or could be textwords denoting methodology in titles and abstracts of articles. We compiled a list of 5,395 terms of which 4,862 were unique and 3,870 returned results (list of terms tested provided by the authors upon request). Examples of the search terms tested are 'disease attributes', 'disease onset', 'early onset', and 'first diagnosis', all as textwords; 'recurrence', the MeSH term, and the MeSH term 'mortality', exploded. The database was randomly split using Microsoft Windows' random number generator into components of 60% and 40%. Search strategies were initially tested and developed in 60% of the database (development) and then validated in 40% of the database (validation).

## Results

Indexing information was downloaded from MEDLINE for 49,028 articles from the 161 journals hand searched. Of these, 1,547 were classified as prognosis, of which 190 (12%) were methodologically sound. Most of the studies classified as prognosis did not assemble an inception cohort and thus 'failed' to be categorized as methodologically sound. Search strategies were developed using all 49,028 articles. Thus the strategies were tested for their ability to retrieve articles about high quality prognostic studies from all other articles, including both low quality prognostic studies and all non-prognostic studies. Table 1 shows the best single term for high-sensitivity, high-specificity, and best balance of sensitivity and specificity from the development database and the operating characteristics of this term in the validation database. The same term, 'exp epidemiologic studies', was identified as the best per-

former in all three areas. When comparing the operating characteristics of 'exp epidemiologic studies' in the development and validation databases, performance was slightly better in the validation database for specificity, precision, and accuracy. For sensitivity, an 8.5% increase was noted in the validation database, but this difference was not statistically significant. A clinical end-user of MEDLINE may find that searching with this single term is worthwhile when using interfaces that do not store the more complex search strategy. This single term is easy to remember and will provide the best retrieval compared with any other single methodologic search term.

Combination of terms with the best results for sensitivity, specificity and optimization of sensitivity and specificity are shown in Table 2. When combining terms to maximize sensitivity while keeping specificity at  $\geq 50\%$ , both sensitivity and specificity were increased. A large increase was achieved for sensitivity – a 25.2% absolute increase – with a much smaller increase of 1.1% achieved for specificity. When terms were combined to maximize specificity while keeping sensitivity at  $\geq 50\%$ , specificity was increased (15.5% absolute increase) but this was done at the expense of sensitivity (decrease of 12.6%). The best optimization of sensitivity and specificity was found with a combination of terms that yielded both sensitivity and specificity at 83%. In most instances the differences in results when comparing the performance in the development and validation databases were nonexistent or very small.

## Discussion

Our study documents search strategies that can help discriminate higher quality from lower quality articles on the prognosis of health disorders. Those interested in all articles reporting high quality studies on prognosis and who are willing to sort out less relevant articles, will probably want to use the most sensitive search strategy. Those with little time to sort through articles and who are looking for a few good articles on prognosis will want to use the most specific strategies. The strategies that optimized

**Table 2: Combination of terms with the best sensitivity (keeping specificity ≥50%), best specificity (keeping sensitivity ≥50%), and best optimization of sensitivity and specificity (based on abs [sensitivity-specificity] < 1%) for detecting studies of prognosis in MEDLINE in 2000**

Search Strategy OVID search*	Sensitivity (%) Development Validation Diff (95% CI) <sup>†</sup>	Specificity (%) Development Validation Diff (95% CI) <sup>†</sup>	Precision (%) Development Validation Diff (95% CI) <sup>†</sup>	Accuracy (%) Development Validation Diff (95% CI) <sup>†</sup>
<b>Best Sensitivity</b>				
incidence.sh.	90.1	79.7	1.7	79.7
OR exp mortality	82.3	79.7	1.6	79.7
OR follow-up studies.sh.	-7.8 (-17.9 to 2.3)	0	-0.1 (-0.5 to 0.5)	0
OR prognos.:tw.				
OR predict.:tw.				
OR course.:tw.				
<b>Best Specificity</b>				
prognos.:tw.	52.3	94.1	3.3	93.9
OR first episode.tw.	48.1	94.2	3.2	94.0
OR cohort.tw.	-4.2 (-18.6 to 10.3)	0.1 (-0.3 to 0.5)	-0.1 (-1.3 to 1.3)	0.1 (-0.4 to 0.5)
<b>Best Optimization of Sensitivity &amp; Specificity</b>				
prognosis.sh.	82.9	83.7	1.9	83.7
OR diagnosed.tw.	73.4	84.1	1.8	84.0
OR cohort.mp.	-9.5 (-21.5 to 2.5)	-0.4 (-0.2 to 1.1)	-0.1 (-0.7 to 0.5)	0.3 (-0.2 to 1.0)
OR predictor.:tw.				
OR death.tw.				
OR exp models, statistical				

\*Search strategies are reported using Ovid's search engine syntax for MEDLINE. The PubMed syntax is embedded in PubMed's Clinical Queries (see Discussion). <sup>†</sup>Diff = Difference, comparing the development and validation data sets using the iterative method of Miettinen and Nurminen for two independent binomial proportions. None of the differences were statistically significant. sh = subject heading; exp = explode, a search term that automatically includes closely related indexing terms; : = truncation; tw = textword (word or phrase appears in title or abstract); mp = multiple posting (term appears in title, abstract, or MeSH heading).

sensitivity and specificity while minimizing the difference between the two provide the best separation of hits from false drops (studies that meet criteria but are not retrieved by the strategy) but do so without regard for whether sensitivity or specificity is affected.

In all cases precision was low. This occurs because of the very low proportion of relevant studies on prognosis in the very large, multipurpose MEDLINE database. Sensitivity and specificity are not affected by the proportion of high quality articles in the database. Precision, on the other hand, is dependent on this proportion, and so is accuracy but to a lesser extent. Low precision means that searchers will continue to need to invest their time in discarding irrelevant retrievals. The low precision values found here should not be over-interpreted: searches were not limited by clinical content terms, as would be the usual case in clinical searches. It may be possible to increase precision and other performance measures by combining search strategies in these tables with methodologic terms using the Boolean 'AND NOT'; by combining search strategies with content specific terms using the Boolean 'AND', for example, 'myocardial infarction AND exp epidemiologic studies'; by multivariate statistical modeling; or by natural language processing. An increase in performance cannot be assumed, however. The next phases of our project will focus on finding better search

strategies through using more sophisticated strategies such as these. We are currently testing the methodologic filters by combining them with disease-specific terms in the discipline areas of mental health and infectious disease as well as the disease-specific area of tuberculosis.

Compared with the performance of search terms for prognosis that we developed in 1991 [11,12], the best performing strategy for sensitivity was the same as that reported in 1991. The operating characteristics for this strategy were similar when comparing the performance in 1991 and 2000 for sensitivity (92% vs 90% [2000]) and specificity (73% vs 80% [2000]) but were quite different for precision (11% vs 2% [2000]). This difference in precision is to be expected given the increased size and diversity of the database in 2000. This shows the robustness of strategies reported in our original study and suggests that search terms in the area of prognosis do not need to be calibrated on large numbers of journals.

The empirical approach that we used for developing search strategies of considering all possible MeSH, publication types, subheadings, and textwords is likely to produce more robust search strategies than any approach based on beginning with a logical MeSH strategy, then adding textwords, subheadings, and publication types.

Those wishing to test their strategies against the ones reported in this paper are invited to send them to us.

The National Library of Medicine (NLM) has updated the Clinical Queries interface of MEDLINE to reflect our new strategies for maximizing sensitivity and maximizing specificity [13]. The translation from Ovid to PubMed syntax was done by staff of the NLM, and compared for performance by the senior author (RBH). SKOLAR MD has also implemented our high specificity strategies [16] and both sensitive and specific strategies have been incorporated into Ovid's main search engine for MEDLINE [17].

## Conclusion

Empirically derived search strategies combining indexing terms and textwords can achieve high sensitivity and specificity for retrieving sound prognostic studies from MEDLINE.

## List of abbreviations

MeSH Medical Subject Headings

exp explosions

pt publication types

sh subject headings

tw textwords

## Competing interests

None declared.

## Authors' contributions

NLW and RBH prepared grant submissions in relation to this project. Both authors drafted, commented on and approved the final manuscript. Both authors also supplied intellect content to the collection and analysis of the data. NLW participated in the data collection and both authors were involved in data analysis and staff supervision.

## Acknowledgments

This research was funded by the National Library of Medicine, USA. The Hedges Team includes Angela Eady, Brian Haynes, Susan Marks, Ann McKibbon, Doug Morgan, Cindy Walker-Dilks, Stephen Walter, Stephen Werre, Nancy Wilczynski, and Sharon Wong.

## References

- Haynes RB, Sackett DL, Tugwell P: **Problems in the handling of clinical and research evidence by medical practitioners.** *Arch Intern Med* 1983, **143**:1971-1975.
- Martinez JL, Licea Serrato J de D, Jimenez R, Grimes RM: **HIV/AIDS practice pattern, knowledge, and education needs among Hispanic clinicians in Texas, USA, and Nuevo Leon, Mexico.** *Rev Panam Salud Publica* 1998, **4**:14-19.
- Covell DG, Uman GC, Manning PR: **Information needs in office practice: are they being met?** *Ann Intern Med* 1985, **103**:596-599.
- Tomlin Z, Humphrey C, Rogers S: **General practitioners' perceptions of effective health care.** *BMJ* 1999, **318**:1532-1535.
- 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.
- Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, Pifer EA: **Obstacles to answering doctors' questions about patient care with evidence: qualitative study.** *BMJ* 2002, **324**:710.
- Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL: **How to keep up with the medical literature: V. Access by personal computer to the medical literature.** *Ann Intern Med* 1986, **105**:810-816.
- Nwosu CR, Khan KS, Chien PF: **A two-term MEDLINE search strategy for identifying randomized trials in obstetrics and gynecology.** *Obstet Gynecol* 1998, **91**:618-622.
- 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**:150-153.
- Bachmann LM, Coray R, Estermann P, Ter Riet G: **Identifying diagnostic studies in MEDLINE: reducing the number needed to read.** *J Am Med Inform Assoc* 2002, **9**:653-658.
- Wilczynski NL, Walker CJ, McKibbon KA, Haynes RB: **Assessment of methodologic search filters in MEDLINE.** *Proc Annu Symp Comput Appl Med Care* 1993:601-605.
- Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC: **Developing optimal search strategies for detecting clinically sound studies in MEDLINE.** *J Am Med Inform Assoc* 1994, **1**:447-458.
- NCBI PubMed Clinical Queries** [<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>]
- Wilczynski NL, Haynes RB: **Robustness of Empirical Search Strategies for Clinical Content in MEDLINE.** *Proc AMIA Symp* 2002:904-908.
- 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):390-393.
- SKOLAR MD** [<http://www.skolar.com>]
- Ovid** [<http://www.ovid.com>]

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1741-7015/2/23/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

