

groups were significantly different from each other, not that biased results were associated with either type of organisation.

Conclusions

Incremental cost effectiveness ratios submitted to NICE's technology appraisals programme by different types of organisation were significantly different from each other. This questions the appropriate methods and processes used by the institute when determining cost effectiveness, and highlights the need for decision makers to have unhindered access to the methods used to produce cost effectiveness estimates.

We thank all the members of the appraisals team at NICE who helped collect the data, Caroline Sabin (professor of epidemiology and medical statistics, University College London) for her advice on the statistical analysis, Martin Buxton (professor of health economics, Brunel University) for his comments on earlier drafts of the manuscript, and the reviewers for their constructive comments.

Contributors: See bmj.com

Funding: None.

Competing interests: AHM and AJF are members of NICE's appraisals team. DF was a member of the appraisals team when this study was undertaken.

Ethical approval: Not required.

- 1 Friedberg M, Saffran B, Stinson TJ, Nelson W, Bennett CL. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA* 1999;15:1453-7.
- 2 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
- 3 Azimi NA, Welch HG. The effectiveness of cost-effectiveness analysis in containing costs. *J Gen Intern Med* 1998;10:664-9.
- 4 Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1997.
- 5 Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 2004;13:437-52.
- 6 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ* 2001;323:1300-3.
- 7 Van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E ratios alongside clinical trials. *Health Econ* 1994;3:309-19.
- 8 Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.

(Accepted 18 October 2004)

doi 10.1136/bmj.38285.482350.82

Optimal search strategies for retrieving systematic reviews from Medline: analytical survey

Victor M Montori, Nancy L Wilczynski, Douglas Morgan, R Brian Haynes, for the Hedges Team

Abstract

Objective To develop optimal search strategies in Medline for retrieving systematic reviews.

Design Analytical survey.

Data sources 161 journals published in 2000 indexed in Medline.

Main outcome measures The sensitivity, specificity, and precision of retrieval of systematic reviews of 4862 unique terms in 782 485 combinations of one to five terms were determined by comparison with a hand search of all articles (the criterion standard) in 161 journals published during 2000 (49 028 articles).

Results Only 753 (1.5%) of the 49 028 articles were systematic reviews. The most sensitive strategy included five terms and had a sensitivity of 99.9% (95% confidence interval 99.6% to 100%) and a specificity of 52% (51.6% to 52.5%). The strategy that best minimised the difference between sensitivity and specificity had a sensitivity of 98% (97% to 99%) and specificity of 90.8% (90.5% to 91.1%). Highest precision for multiterm strategies, 57% (54% to 60%), was achieved at a sensitivity of 71% (68% to 74%). The term "cochrane database of systematic reviews.jn." was the most precise single term search strategy (sensitivity of 56% (52% to 60%) and precision of 96% (94% to 98%)). These strategies are available through the "limit" screen of Ovid's search interface for Medline.

Conclusions Systematic reviews can be retrieved from Medline with close to perfect sensitivity or specificity, or with high precision, by using empirical search strategies.

Introduction

Finding systematic reviews in Medline poses two challenges. Firstly, only a tiny proportion of citations in Medline are for literature reviews, and only a fraction of these are systematic reviews. Secondly, the National Library of Medicine's Medlars indexing procedures do not include "systematic review" as a "publication type." Rather, the indexing terms and publication types include a number of variants for reviews, including "meta-analysis" (whether or not from a systematic review); "review, academic"; "review, tutorial"; "review literature"; as well as separate terms for articles that often include reviews, such as "consensus development conference", "guideline", and "practice guideline". The need for special search strategies (hedges) for systematic reviews could be substantially reduced if such reviews were indexed by a separate publication type, but indexers need to be able to dependably distinguish systematic reviews from other reviews. Pending this innovation, there is need for validated search strategies for systematic reviews that optimise retrieval for clinical users and researchers.

In this paper we report on the generation, validation, and performance characteristics of new search strategies to identify systematic reviews in Medline, and compare them with previously published strategies.

 A table showing PubMed translations of Ovid search strategies is on bmj.com

 This is the abridged version of an article that was posted on bmj.com on 24 December 2004: <http://bmj.com/cgi/doi/10.1136/bmj.38336.804167.47>

Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA

Victor M Montori
assistant professor

Health Information Research Unit, Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada L8N 3J5

Nancy L Wilczynski
doctoral student

Douglas Morgan
data analyst

R Brian Haynes
professor

Correspondence to: R B Haynes, Department of Clinical Epidemiology and Biostatistics, Room 2C10B Health Sciences Center, McMaster University Faculty of Health Sciences, 1200 Main Street West, Hamilton, ON, Canada L8N 3J5
bhaynes@mcmaster.ca

BMJ 2005;330:68-71

Table 1 Best multiple-term strategies maximising sensitivity and minimising the difference between sensitivity and specificity. Values are percentages (95% confidence intervals)

Search strategy in Ovid format	Sensitivity*	Specificity**	Precision†
Top sensitivity strategies‡			
search:.tw. or meta-analysis.mp.pt. or review.pt. or di.xs. or associated.tw.			
Development	100 (97.3 to 100)	63.5 (62.5 to 64.4)	3.41 (2.86 to 4.03)
Validation without CDSR	99.7 (99.1 to 100)	51.1 (50.7 to 51.6)	1.4 (1.2 to 1.5)
Validation	99.9 (99.6 to 100)	52.0 (51.6 to 52.5)	3.14 (2.92 to 3.37)
Top strategy minimising the difference between sensitivity and specificity§			
meta-analysis.mp.pt. or review.pt or search:.tw.			
Development	92.5 (86.6 to 96.3)	93.0 (92.5 to 93.5)	14.6 (12.3 to 17.2)
Validation without CDSR	95.5 (93.3 to 97.7)	89.9 (89.7 to 90.2)	6.1 (5.5 to 6.8)
Validation	98.0 (97.0 to 99.0)	90.8 (90.5 to 91.1)	14.2 (13.3 to 15.2)

CDSR=Cochrane Database of Systematic Reviews.

*Development dataset (n=133); validation dataset without CDSR (n=332); validation dataset (n=753).

**Development dataset (n=10 313); validation dataset without CDSR (n=48 258); validation dataset (n=48 275).

†Numbers vary by row.

‡Keeping specificity ≥50%; adding the Cochrane Database of Systematic Reviews (using boolean OR) did not improve performance.

§Keeping sensitivity ≥90%.

Methods

We developed search strategies by using methodological search terms and phrases in a subset of Medline records matched with a hand search of the contents of 161 journal titles for 2000; this manual review of the literature represented the criterion standard dataset (see bmj.com).²

Study classification

We defined a review as any full text article that was shown as a review, overview, or meta-analysis in the title or in a section heading, or that indicated in the text that the intention of the authors was to review or summarise the literature on a particular topic.³ For an article to be considered a systematic review, the authors had to clearly state the clinical topic of the review and how the evidence was retrieved and from what sources, and they had to provide explicit inclusion and exclusion criteria and include at least one study that passed methodological criteria for the purpose category. For example, reviews of interventions had to have at least one study with random allocation of participants to comparison groups and assessment of at least one clinical outcome.

Search terms

To construct a comprehensive set of possible search terms, we listed indexing terms (for example, subject headings and subheadings, publication types) and text

words used to describe systematic reviews (single words or phrases that may appear in titles or abstracts, both in full and in various truncations). We sought further terms from clinicians and librarians, and from published strategies from other groups. We compiled a list of 4862 unique terms and tested them using the Ovid Technologies searching system.

Building search strategies

We determined the sensitivity, specificity, and precision of single term and multiple term search strategies against the criterion standard dataset. Sensitivity for a given topic is defined as the proportion of systematic reviews for that topic that are retrieved; specificity is the proportion of non-systematic reviews not retrieved; and precision is the proportion of retrieved articles that are systematic reviews.

Individual search terms with a sensitivity of more than 50% (to develop strategies that optimised sensitivity) and a specificity of more than 75% (to develop strategies that optimised specificity) for identifying systematic reviews were incorporated into the development of two term strategies; we continued this process to build five term strategies that optimised either sensitivity or specificity.

Performing search strategies

We iteratively tested these strategies against a validation dataset—a set of articles from 10 journals in

Table 2 Best multiple-term strategies maximising precision. Values are percentages (95% confidence intervals)

Search strategy in Ovid format	Sensitivity*	Specificity**	Precision†
Top precision performer‡			
Medline.tw. or systematic review.tw. or meta-analysis.pt.			
	75.2 (67.0 to 82.3)	99.4 (99.2 to 99.5)	60.2 (52.4 to 67.7)
	74.4 (69.7 to 79.1)	98.6 (98.5 to 98.7)	26.3 (23.5 to 29.1)
	71.2 (68.0 to 74.4)	99.2 (99.1 to 99.3)	57.1 (53.9 to 60.3)
Combining most precise term with most sensitive terms			
Cochrane database of systematic reviews.jn.§ or search:.tw. or meta-analysis.pt. or Medline.tw. or systematic review.tw.			
	90.2 (88.1 to 92.3)	98.4 (98.3 to 98.5)	46.5 (43.9 to 49.0)
Cochrane database of systematic reviews.jn. or search:.tw.			
	79.0 (76.1 to 81.9)	98.8 (98.8 to 98.9)	51.8 (48.9 to 54.7)
Cochrane database of systematic reviews.jn. or Medline.tw.			
	74.4 (71.3 to 77.5)	99.5 (99.4 to 99.5)	68.8 (65.6 to 72.0)

CDSR=Cochrane Database of Systematic Reviews.

*Development dataset (n=133); validation dataset without CDSR (n=332); validation dataset (n=753).

**Development dataset (n=10 313); validation dataset without CDSR (n=48 258); validation dataset (n=48 275).

†Numbers vary by row.

‡Keeping sensitivity ≥75%.

§The term for the Cochrane Database of Systematic Reviews (Cochrane database of systematic reviews.jn.) does not apply to the development or to the validation without—CDSR dataset—since it does not include records from that source.

Table 3 Performance from published strategies to identify systematic reviews in Medline tested in our full validation database. Values are percentages (95% confidence intervals)

Search strategy	Sensitivity*	Specificity**	Precision†
Centre for Reviews and Dissemination			
High sensitivity (16 terms)	97.6 (96.5 to 98.7)	69.6 (69.2 to 70.0)	4.77 (4.43 to 5.11)
Intermediate sensitivity and precision (29 terms)	96.7 (95.4 to 98.0)	79.7 (79.3 to 80.0)	6.91 (6.42 to 7.39)
High sensitivity and precision (12 terms)	95.8 (94.3 to 97.2)	89.7 (89.4 to 90.0)	12.7 (11.8 to 13.5)
Hunt and McKibbin			
Simple query (4 terms)	68.8 (65.5 to 72.1)	99.2 (99.1 to 99.3)	56.7 (53.5 to 59.9)
Sensitive query (8 terms)	73.4 (70.3 to 76.6)	99.1 (99.0 to 99.2)	55.1 (52.0 to 58.2)
Shojania and Bero			
PubMed based query (71 terms)	90.0 (87.9 to 92.2)	97.2 (97.0 to 97.4)	33.2 (31.2 to 35.2)
Hedges (this report)			
Sensitive query (5 terms)‡	99.9 (99.6 to 100)	52.0 (51.6 to 52.5)	3.14 (2.92 to 3.37)
Balanced query, sensitivity>specificity (3 terms)§	98.0 (97.0 to 99.0)	90.8 (90.5 to 91.1)	14.2 (13.3 to 15.2)
Balanced query, specificity>sensitivity (5 terms)¶	90.2 (88.1 to 92.3)	98.4 (98.3 to 98.5)	46.5 (43.9 to 49.0)
Specific query (3 terms)††	71.2 (68.0 to 74.4)	99.2 (99.1 to 99.3)	57.1 (53.9 to 60.3)

*Validation (n=753).

**Validation (n=48 275).

†Numbers vary by row.

‡search.tw. or meta-analysis.mp.pt. or review.pt. or di.xs. or associated.tw.

§ meta-analysis.mp.pt. or review.pt. or search.tw.

¶Cochrane database of systematic reviews.jn. or search.tw. or meta-analysis.pt. or Medline.tw. or systematic review.tw.

††Medline.tw. or systematic review.tw. or meta-analysis.pt.

which the highest proportion of systematic reviews are published² without including the Cochrane Database of Systematic Reviews. We sorted these strings by sensitivity, by specificity, by the absolute difference of sensitivity and specificity, and by precision. Then we selected representative strings of terms among those with similar top performance (within 3% of each other).

To validate the strategies, we tested candidate strategies in the two validation datasets: database comprising records from 161 journals, one with and one without the Cochrane Database of Systematic Reviews.

Results

The derivation database had 10 446 records, of which 133 (1.3%) were systematic reviews. The full validation database (including the Cochrane Database of Systematic Reviews) included 49 028 records, of which 753 (1.7%) were systematic reviews. The single term strategies that performed best are shown on [bmj.com](#). Table 1 shows the top strategies that maximise sensitivity and minimise the absolute difference between sensitivity and specificity (while keeping both $\geq 90\%$), a strategy that optimises the balance of sensitivity and specificity. Table 2 shows a strategy with top precision and a set of strategies that identify systematic reviews with greater sensitivity and precision resulting from combining the term “cochrane database of systematic reviews.jn.”, a top precision performer, with each of the terms that performed best.

Table 3 describes the performance of the most popular strategies available to search for systematic reviews when tested against our full validation database. With the exception of Hunt and McKibbin strategies, the five term balanced query and the three term specific query offer higher specificity and precision than the other strategies.

Discussion

Our study documents search terms with best sensitivity, best specificity and precision, and smallest

difference between sensitivity and specificity for retrieving systematic reviews from Medline. As a result, we offer Medline hedges with fewer terms and better performance than the best available strategies. These strategies can be implemented using the “limit” top button on the Ovid’s search interface for Medline ([www.ovid.com](#)) or users can “copy and paste” the strategies in table A (see [bmj.com](#)) into a PubMed query ([www.pubmed.gov](#)), with the chosen strategy combined (by using boolean AND) with appropriate content terms. Also, the top strategy minimising the difference between sensitivity and specificity has been incorporated into Skolar ([www.skolar.com](#)).

Implications for future research and users of medical literature

Ideally, systematic reviews should be indexed using an exact publication type term. Unfortunately, accurate application of this term requires judgment based on assessment of the methods reported in the original articles being indexed. Our process for doing this is highly reproducible, but it has not been shown whether indexers can be trained to do this, and to do so in the time that they have for applying index terms. Pending this, users of the medical literature will need to use hedges, such as those offered here, to identify systematic reviews in Medline.

What is already known on this topic

Systematic reviews are important for advancing science and evidence based clinical practice, but they may be difficult to retrieve from Medline

What this study adds

Special search strategies retrieved up to 99.9% of systematic reviews or were able to maximise the proportion of citations retrieved that are systematic reviews

Currently, clinicians can search systematic reviews within the Cochrane Library, where they can find the Cochrane Database of Systematic Reviews and DARE. If they cannot find a pertinent review, or their interest is other than prevention and treatment, or if they want to conduct a comprehensive search, they can use the strategies presented here to identify systematic reviews in Medline. Quick searches or searching for systematic reviews in topic areas where many are available may be optimal with a high precision strategy or with the strategy that balanced sensitivity and precision. On the other hand, guideline developers and researchers may want to use a highly sensitive strategy. For all, our strategies are most useful when they are preprogrammed into search interfaces, such as the Clinical Queries in PubMed, ready to be combined with topic specific terms.

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

Contributors: See bmj.com

Funding: This research was funded by the National Library of Medicine, USA (grant No 1 RO1 LM06866). VMM is a Mayo Foundation Scholar. These funding sources had no additional role.

Competing interests: None declared.

Ethical approval: Not required.

- 1 Dickersin K, Higgins K, Meinert CL. Identification of meta-analyses. The need for standard terminology. *Control Clin Trials* 1990;11:52-66.
- 2 Montori VM, Wilczynski NL, Morgan D, Haynes RB. Systematic reviews: a cross-sectional study of location and citation counts. *BMC Med* 2003;1:2.
- 3 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;10:390-3. (Accepted 4 October 2004)

doi 10.1136/bmj.38336.804167.47

Rise in “no indicated risk” primary caesareans in the United States, 1991-2001: cross sectional analysis

Eugene Declercq, Fay Menacker, Marian MacDorman

This paper analyses US national birth certificate data on approximately 4 million births annually to create a new category—mothers at “no indicated risk”—and then examines the growth of primary caesareans in these women from 1991 to 2001. No indicated risk denotes mothers with singleton, full term (≥ 37 weeks), vertex presentation births who were not reported to have any medical risk factors and for whom no complications of labour or delivery were listed on the birth certificate. (See bmj.com for definitions.)

Methods and results

The proportion of mothers at no indicated risk decreased from 46% of all births in 1991 to 42% in 1998 but has since levelled off (table). However, the primary caesarean rate for this exceptionally low risk group rose 67% between 1991 (3.3%) and 2001 (5.5%), with a gradual increase from 1991 to 1996 and a rapid one thereafter.

Older, primiparous mothers were much more likely to have a no indicated risk primary caesarean; almost one fifth (19.5%) of primiparous mothers aged over 34 had such a delivery in 2001. More than 5% of multiparous mothers over 34 who had had previous vaginal births also had a no indicated risk primary caesarean in 2001. Among mothers under 30 with no indicated risk, the primary caesarean rate grew by more than half (58%) between 1991 and 2001 to 4.9%.

The raw numbers of births also illustrates this trend. In 2001, 80 028 no indicated risk primary caesareans took place in the United States, an increase of 25 162 since 1996. This represented 25.8% of the total increase (97 659) in primary caesareans between 1996 and 2001.

We used multivariate logistic regression analysis (SAS version 8) to examine changes in primary caesarean rates after controlling for parity; maternal ethnicity, age, and education; birth weight; and data year (1991, 1996, or 2001) (see table on bmj.com). We ran models for all mothers, including parity as a variable, and for first time mothers only. Age was a major factor, particularly among first time mothers. For primiparous mothers aged over 40, the odds of having a caesarean were 5.4 times that for mothers aged 20-24. In the multivariate analysis, the overall increase between 1991 and 1996 disappeared, but the odds of having a no indicated risk primary caesarean in 2001 were almost 50% higher than the odds for comparable mothers in 1996.

Maternal and Child Health Department, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118-2526, USA

Eugene Declercq
professor, maternal and child health

Division of Vital Statistics, National Center for Health Statistics, Hyattsville, MD 20782, USA

Fay Menacker
statistician

Marian MacDorman
statistician

Correspondence to: E Declercq
declercq@bu.edu

BMJ 2005;330:71-2

What is already known on this topic

The overall and primary caesarean rate is growing rapidly in the United States and worldwide, and the likelihood of a caesarean is strongly related to age of the mother and parity

What this study adds

A new category for analysis has been created—the “no indicated risk” caesarean

The proportion of no indicated risk primary caesareans is growing rapidly in the United States, adding to the overall rise in the primary caesarean rate



Definitions and an extra table are on bmj.com

This article was posted on bmj.com on 19 November 2004: <http://bmj.com/cgi/doi/10.1136/bmj.38279.705336.0B>