Systematic Reviews

Identifying relevant studies for systematic reviews

Kay Dickersin, Roberta Scherer, Carol Lefebvre
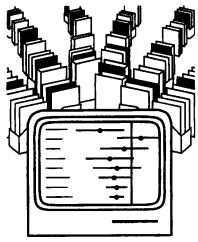
Abstract

Objective—To examine the sensitivity and precision of Medline searching for randomised clinical trials.

Design—Comparison of results of Medline searches to a "gold standard" of known randomised clinical trials in ophthalmology published in 1988; systematic review (meta-analysis) of results of similar, but separate, studies from many fields of medicine.

Populations—Randomised clinical trials published in 1988 in journals indexed in Medline, and those not indexed in Medline and identified by hand search, comprised the gold standard. Gold standards for the other studies combined in the meta-analysis were based on: randomised clinical trials published in any journal, whether indexed in Medline or not; those published in any journal indexed in Medline; or those published in a selected group of journals indexed in Medline.

Main outcome measure—Sensitivity (proportion of the total number of known randomised clinical trials identified by the search) and precision (proportion of publications retrieved by Medline that were actually randomised clinical trials) were calculated for each study and combined to obtain weighted means. Searches producing the "best" sensitivity were used for sensitivity and precision estimates when multiple searches were performed.

Results—The sensitivity of searching for ophthalmology randomised clinical trials published in 1988 was 82%, when the gold standard was for any journal, 87% for any journal indexed in Medline, and 88% for selected journals indexed in Medline. Weighted means for sensitivity across all studies were 51%, 77%, and 63%, respectively. The weighted mean for precision was 8% (median 32·5%). Most searchers seemed not to use freetext subject terms and truncation of those terms.

Conclusion—Although the indexing terms available for searching Medline for randomised clinical trials have improved, sensitivity still remains unsatisfactory. A mechanism is needed to "register" known trials, preferably by retrospective tagging of Medline entries, and incorporating trials published before 1966 and in journals not indexed by Medline into the system.

Background

Of primary importance to the results obtained in a systematic review or meta-analysis are the data collection methods used. Data collection includes all methods used to identify published and unpublished data to be included in the review, to determine eligibility of the data for inclusion, and to extract data for analysis. Much of the methodological research related to meta-analysis has, to date, related to statistical methods, not to data collection.[1] It is clear, however, that the validity of the results of statistical analyses depends on the validity of the underlying data.

Unbiased and complete identification of studies is particularly important. Studies relevant for inclusion in a systematic review may not have been published, for reasons related to the findings (publication bias).[2-5] Even when studies are published, they may be difficult to find.

Our objective in the studies we describe was to examine the sensitivity and precision of searching Medline for randomised clinical trials. In this setting, sensitivity is defined as the proportion of the total number of known trials identified by the search and precision is the proportion of publications retrieved by Medline that are actually randomised clinical trials. We performed a two part study, the first specifically searching for randomised clinical trials in vision research (ophthalmology and optometry) and the second combining the results from the vision study with other similar studies to achieve a better estimate of sensitivity and precision of searching across medical disciplines.

Methods

SEARCHING FOR RANDOMISED CLINICAL TRIALS ON VISION

In 1991, in collaboration with the health sciences library at the University of Maryland at Baltimore, we developed a two stage search strategy designed to identify randomised clinical trials in vision research. A clinical trial was defined (by C Meinert, 1991) as any planned therapeutic, diagnostic, or preventive study involving humans comparing concurrently one intervention (drug, device, or procedure) to another intervention, placebo, or no intervention to determine their relative safety and efficacy. Randomised trials were those in which treatment was truly randomised by using a computer generated list or a random numbers table or in which a quasi-randomisation method, such as assignment to treatment by medical record number, was used. Our overall goal was to develop a register of published trials in vision research. The activities described in this article were related to a pilot study for developing the register; the study was designed to devise the best possible strategy for identifying as high a proportion as possible of published randomised clinical trials.

The first search of Medline for reports of randomised clinical trials published in 1988 was relatively broad both in terms of the subject matter covered and the methodological MeSH (medical subject heading) terms (details of strategies used in this study are available from corresponding author). Citations and abstracts retrieved by the first search were reviewed, and potentially relevant trials were downloaded into Pro-Cite version 1.41 using Biblio-Links.

To determine the "gold standard," all journals that had appeared at least once in the first Medline search results (44 journals) were selected for hand searching for articles reporting randomised clinical trials of research on vision. Thirty nine of the 44 journals were available in local libraries. An additional 27 ophthalmology and optometry journals were selected for hand searching because they included English abstracts and were available in local libraries. Fourteen of these journals were not indexed in Medline. Thus a total of 66 journals were hand searched and reports of

Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA
Kay Dickersin, *assistant professor*
Roberta Scherer, *project coordinator*

UK Cochrane Centre, Oxford OX2 7LG
Carol Lefebvre, *information specialist*

Correspondence to:
Dr Dickersin.

randomised clinical trials photocopied for our files. When it was not clear whether an article described a randomised trial, it was given the benefit of the doubt and included in the initial set of articles we used to refine our search strategy.

In the second part of the vision study we performed an analysis of the text words contained in the title and abstract and the MeSH terms used to index the reports found by the first Medline search and hand searching. This information was used to devise a second Medline search strategy, designed to be more sensitive than the first. The list of references retrieved was reviewed, and new trials were identified and their reports retrieved from the library. These reports plus those identified initially were reviewed by two of us (RS and KD) independently for inclusion in the gold standard.

We further tested the second Medline search strategy by applying it to articles published in 1989. Subsequently, we hand searched for 1989 the four journals publishing the greatest number of randomised clinical trials on vision in our 1988 gold standard.

Additional information and analysis

There were two situations in which it was not clear whether an article described a randomised trial. In the first, when articles were published in languages other than those we read, we had the article translated to the degree necessary to determine whether it met our definition of a randomised clinical trial. In the second, when it was not clear from the written report whether a random (or quasi-random) method had been used to allocate patients to an intervention, addresses were obtained and letters requesting this information were written to the first authors of the reports. Second and succeeding authors were written to if no address was available for the first author. If the author stated that the trial used a specific method to randomly assign patients to treatment the report was included in the gold standard. We made no attempt to verify statements in published reports.

We compared the 1988 results of our first and second Medline searches with the gold standard to calculate the sensitivity and precision of the searches. For 1989 reports, we compared the results of the search with the 1989 gold standard compiled by hand searching the four journals. Preliminary results of our search[6 7] have here been updated subsequent to translation of articles and correspondence with authors.

SYSTEMATIC REVIEW
Identification of studies

We included only articles that reported results of searches for randomised clinical trials in the methods section or when confirmed by correspondence with the author(s). Our focus on trials is related to our interest in systematic reviews of trials evaluating selected interventions. Several published studies on the sensitivity of Medline searching which were not limited to searching for randomised clinical trials were therefore not included.

We identified studies for our review by using both formal and ad hoc methods. Because of our interest in the problem of Medline searching, one of us (KD) has conducted Medline searches over the years to identify relevant articles. These search strategies have not been standardised; rather, they have depended on the use of text words such as "Medline" and "searching." Bibliographies of these and other related articles have been reviewed for pertinent reports. For this review, CL performed a formal search of Medline and Embase for articles not previously identified. Also, through an international meeting of investigators who conducted original research in this area, convened in November 1992 at the UK Cochrane Centre in Oxford, we learned of new publications not identified using the other methods described.

Letters were written to authors of studies when additional data or clarifications were needed for our analyses. Thus, some information relating to studies in this systematic review may not have been published previously or may differ slightly from that in the published article.

Gold standard

The studies included in the review varied somewhat in their approach, but all compared the results of a Medline search to a gold standard of known, published randomised trials. There were three types of gold standards based on known published trials: trial reports published in journals, books, or proceedings, including publications not indexed by Medline; trial reports published in journals indexed in Medline; and trial reports published in selected Medline journals. For example, the searches for randomised clinical trials of intraventricular haemorrhage and neonatal hyperbilirubinaemia[8] used the Oxford Database of Perinatal Trials, which included trial reports from a variety of sources as the gold standard. On the other hand, searches for pain trials[9] focused on seven selected journals covered by Medline; this gold standard was developed by hand searching these journals back to 1966.

In several cases authors of individual studies excluded articles from the numerator (number of studies identified) of the sensitivity and precision calculations because certain criteria for the review were not met. For example, criteria in Dickersin *et al* required articles in English, published between 1966 and 1983.[8] In such cases, articles not meeting the study criteria were also excluded from the denominator (either the gold standard or the number of citations identified by a search). Two studies reported cases where articles should have been included in the Medline database, by virtue of being published in a journal indexed for Medline, but were not.[10 11] In such cases, these articles were allowed to remain in the Medline gold standards, even though they were not in the Medline file.

Analyses

Analyses compared the sensitivity and precision of results of the Medline search for the individual studies and combined these findings by adding numerators and denominators to obtain weighted means. When an article described the results of multiple search strategies we used the results from the strategy providing the highest sensitivity when we calculated sensitivity. Similarly, when calculating precision, we used figures corresponding to the search providing the highest sensitivity. In two cases, a second group of investigators replicated a search and came up with a strategy that improved the sensitivity of the search (Poynard and Conn[12] and Bernstein[13]; Silagy[14] and Jadad and McQuay[15]); we reported only the higher sensitivity, where the denominators (gold standards) were identical. Since there was 100% concordance in the gold standards used by Silagy and Jadad and McQuay, and a single sensitivity was extracted for our meta-analysis, we considered these articles to represent a single "study."

We also explored the individual search strategies in an effort to understand differences in retrieval rates. Possible reasons for less than optimal sensitivity were classified into five broad categories: limited use of subject matter MeSH terms; limited use of methodological controlled vocabulary (MeSH, check tags, and publication types); limited use of freetext subject matter terms; limited use of freetext methodological terms; and limited use of truncation.

Results

SEARCHING FOR RANDOMISED CLINICAL TRIALS ON VISION

The gold standard for randomised clinical trials on vision published in 1988, developed by using a combination of hand search and Medline search, comprised a total of 236 reports classified as randomised clinical trials (table I). Forty eight of the reports (20%) were in languages other than English; 222 (94%) were published in Medline journals and 14 (6%) elsewhere.

The first Medline search for articles published in 1988 resulted in 219 references, of which 105 were classified as randomised clinical trials (table II)—a precision of 48%. The sensitivity of the first Medline search was 44% (105/236) in comparison with the gold standard that included all randomised clinical trials published in any journal, and 47% (105/222) for trials reported in Medline journals.

The second Medline search resulted in 1520 references, of which 193 were identified as randomised clinical trials (precision=13%). Sensitivity of the search using all known randomised clinical trials as the gold standard was 82% and sensitivity using only trials listed in Medline was 87%. The second search identified eight trials missed by both the first Medline search and the hand searches, as well as 24 trials appearing in 17 journals not originally hand searched. A hand search of four of these journals (those that contained more than one citation) resulted in no further additions.

The results of the Medline search for articles

TABLE I—*"Gold standard" for randomised clinical trials in vision research published in 1988: number (percentage) of reports of randomised clinical trials in Medline and non-Medline journals by language*

Language	Medline journals	Non-Medline journals	Total
English	183 (82)	5 (36)	188 (80)
Other	39 (18)	9 (64)	48 (20)
Total	222 (100)	14 (100)	236 (100)

TABLE II—*Sensitivity and precision of Medline searches for randomised clinical trials of research on vision*

	No of citations	No of randomised clinical trials found	Precision (%)	Sensitivity (%) Overall (n=236)	Sensitivity (%) Medline journals (n=222)	Sensitivity (%) Selected journals (n=61)
1988:						
Search 1	219	105	48	44	47	
Search 2	1520	193	13	82	87	
1989*:						
Search 2	272	54	20			88

*Four journals searched.

TABLE III—*"Best" sensitivity of Medline searches**

Year	First author	Topic	Proportion (of "gold standard") identified (%) Any publication	Proportion (of "gold standard") identified (%) Any Medline journal	Proportion (of "gold standard") identified (%) Select Medline journals
1985	Dickersin	Intraventricular haemorrhage	(19/34) 56	(19/32) 59	
		Hyperbilirubinaemia	(28/96) 29	(28/88) 32	
1985	Poynard**	Liver disease	(107/279) 38		
1988	Bernstein**	Liver disease		(155/195) 79	
1989	Ohlsson	Pregnancy	(10/13) 77	(10/11) 91	
1990	Hofmans	Acupuncture	(57/98) 58	(56/67) 85	
1991	Gøtzsche	Rheumatoid arthritis	(128/200) 64	(128/140) 91	
1992	Kleijnen	Homoeopathy	(18/107) 17	(18/23) 78	
		Vitamin C	(22/61) 36	(22/28) 78	
		Ginkgo	(14/45) 31	(14/18) 78	
1994	Dickersin	Ophthalmology	(193/236) 82	(193/222) 87	
Unpublished	Lacy	Newborn	(8/17) 47	(8/17) 47	
1989	Kirpalani	Newborn			(28/53) 53
1993	Jadad	Pain			(126/153) 82
1993	Jadad/Silagy	Primary care			(179/204) 88
1994	Dickersin	Ophthalmology			(54/61) 88
1994	Adams	Mental health			(388/746) 52
1994	Solomon	Surgery			(17/37) 46
(Totals) Weighted means			(604/1186) 51	(651/841) 77	(792/1254) 63
Range			17-82	32-91	46-88

*Some numbers may vary from publication because clarified by communication with author.
**Same search team and subject.

published in the four selected journals in 1989, using the second strategy, were similar to those for 1988 in terms of sensitivity. Sixty one reports classified as randomised clinical trials and published in one of the four journals comprised the gold standard. The Medline search retrieved 272 citations for the four journals, of which 54 were confirmed as randomised clinical trials (precision=20%, sensitivity=88%).

SYSTEMATIC REVIEW

For our review, we identified 12 relevant articles published in journals,[8-10 12-20] data from three studies not yet published at the time of our review,[11 12 22] and data from the first part of our study, reported above (referred to in tables III-V as Dickersin 1994). In each case, investigators performed Medline searches and compared the results of the searches with various gold standards of known trials.

Of the 16 studies identified that examined Medline searching for randomised clinical trials, we obtained information useful for this review from 15. On average, the studies indicated that a Medline search, even when conducted by a trained searcher, yielded only 51% of all known trials (range 17-82%; table III). With a gold standard of only those trials in journals available on Medline, sensitivity was better but still disappointing, at 77%, and for studies that used specially selected Medline journals as a gold standard the weighted average sensitivity was 63% (46-88%).

Some of the studies that investigated the sensitivity of Medline searching also examined strategies that would maximise sensitivity while minimising the number of citations that would have to be reviewed for potential relevance. The results of these studies show that there is wide variation in the "precision" that can be achieved in searching. For some topics, thousands of citations must be examined to achieve acceptable sensitivity; for others, a relatively small number of citations require review (median precision 32·5% (2-82%); table IV).

The differences in the sensitivities achieved with searching may be due to earlier studies having limited their use of subject matter MeSH terms too severely, either in the area of subheadings[12] or other related terms[18 19] (table V). Other common deficiencies are limited use of free text terms and limited use of truncated text terms.

Discussion

The sensitivity of Medline is 51% when the gold standard is all known randomised clinical trials published in journals indexed in Medline and in those not indexed in Medline. Thus, if comprehensive systematic reviews of randomised clinical trials depend solely on Medline searches they will omit about half of the available studies. It is not even possible to identify all the published trials in journals indexed in Medline by using Medline (weighted mean=76%).

There are about 22 000 active medical serial titles,[23] of which about 16 000 can be classified as journals; only about 3700 of these are in Medline. Not all 16 000 journals are likely to publish the reports of randomised trials, but many report results of randomised clinical trials presented at meetings, only half of which ever reach full publication.[24] It might be argued that the quality of reports in non-Medline journals is lower than that of reports in Medline journals and thus missing randomised clinical trials reported in non-Medline journals in a systematic review might be relatively unimportant—but there is no evidence that this is so.

INADEQUATE INDEXING

Sensitivity was much better, about 77%, on comparison of the results of Medline searching only with a

Year	Author	Topic	No of relevant trials	No retrieved	%Of relevant trials
		Any Medline journal			
1985	Dickersin	Intraventricular haemorrhage	19	36	53
		Hyperbilirubinaemia	28	39	72
1988	Bernstein	Liver disease	155	9 643	2
1989	Ohlsson	Pregnancy	10	125	8
1991	Gøtzsche	Rheumatoid arthritis	128	738	17
1992	Kleijnen	Homoeopathy	18	52	35
		Vitamin C	22	81	27
		Ginkgo	14	46	30
1994	Dickersin	Ophthalmology	193	1 520	13
Unpublished	Lacy	Newborn	8	21	38
Total (weighted mean; median)			595	12 301	(5; 29)
		Select Medline journals			
1989	Kirpalani	Newborn	28	34	82
1993	Jadad	Pain	126	245	51
1994	Dickersin	Ophthalmology	54	272	20
1994	Adams	Mental health	388	662	59
Total (weighted mean; median)			596	1 213	(49; 55)
Grand total (weighted mean; median)			1 191	13 514	(9; 33)
Range					(2%-82%)

TABLE V—*Possible reasons for non-optimal sensitivity of "best" Medline searches*

Year	First author	Limited use of MeSH terms		Limited use of free text terms		Limited use of truncated text terms
		Subject matter	Methods	Subject matter	Methods	
1985	Dickersin					
1985	Poynard	✓	✓	✓	✓	✓
1988	Bernstein			✓		✓
1989	Kirpalani	✓		✓		✓
1989	Ohlsson	✓		✓		✓
1990	Hofmans	✓	✓	✓	✓	✓
1991	Gøtzsche			✓		✓
1992	Kleijnen			✓	✓	✓
1993	Jadad					✓
1993	Jadad/Silagy					
1994	Dickersin					✓
1994	Adams (in press)	*		*		
Unpublished	Lacy		No data	✓	No data	✓
1994	Solomon			✓		

*Adams *et al* did not use any "subject" terms in their search strategy. Instead they used journal title to restrict their retrieval.

gold standard of known randomised clinical trials that are included in the Medline file. This proportion could and should be 100%; the problem results mainly from inadequate indexing, for which there are several reasons. Firstly, until fairly recently there has been an emphasis on developing MeSH terms for subject matter rather than methodology. For example, there was no suitable descriptor term to describe randomisation as a methodology until RANDOM ALLOCATION was introduced in 1978. RANDOMIZED CONTROLLED TRIALS was introduced in 1990 as a descriptor term and RANDOMIZED CONTROLLED TRIAL was introduced in 1992 as a publication type. Secondly, even when suitable descriptor terms were available, they have not been and continue not to be applied consistently by indexers acting for the National Library of Medicine (P L Schuyler *et al*, second international congress on peer review, 1993; C Lefebvre *et al*, conference on "An evidence-based health care system: the case for clinical
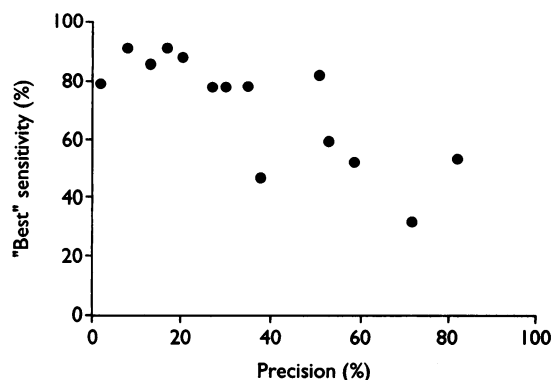
Relation between "best" sensitivity (gold standard obtained for either any journal indexed by Medline or selected Medline journals) and precision for Medline searching of 14 topics

trial registries" at National Institutes of Health, 1993). Thirdly, authors may not have described their research methods clearly enough to allow accurate indexing of methodology. For example, 11% (25/236) of the trials comprising the broadest gold standard in any publication of vision research could not be verified as randomised clinical trials by readers and required confirmation by a letter to the author. For an additional 37 articles, the authors did not respond to inquiries or could not be reached. These articles may have been randomised clinical trials but could not be included in the gold standard because their status remains unclear.

FACTORS AFFECTING SENSITIVITY AND PRECISION

The calculation of sensitivity requires comparing the results of a Medline search with a gold standard of known randomised clinical trials. Two major factors will influence this calculation. The first is the comprehensiveness of the gold standard. It is likely that the more comprehensive the gold standard, the less sensitive the Medline search, particularly if the gold standard includes many randomised clinical trials in journals not indexed in Medline. Thus, differences across studies in sensitivity may be related to the completeness of the gold standard or of the field itself. We examined the available data by using three possible gold standards, so the sensitivities presented therefore address different questions. The sensitivity of searches using trial reports from any publication as a denominator expresses, at least theoretically, the probability of identifying all published randomised clinical trials in a field, although few gold standards are likely to be complete. The sensitivity of searches using reports from any Medline journal as a denominator expresses the probability of identifying randomised clinical trials known to be available on Medline. Assuming the investigators have done a thorough job, this gold standard is more likely to be "complete" because the universe of trials indexed in Medline is well defined. The gold standard using selected Medline journals can be easily checked for reliability and validity because it uses a specific subset of journals published within a defined time period. If the journals included in this gold standard were representative of all journals over all time periods it would determine an overall sensitivity of Medline searching. It is not likely to be representative, however, and thus its chief value is that the denominator of the sensitivity calculation (the gold standard) is likely to be accurate.

The second factor influencing the calculation of sensitivity and precision estimates is the quality of the Medline search. Because Medline is a highly structured database with complex indexing rules, a certain level of skill and experience is necessary to achieve good (sensitive and precise) results. Untrained Medline searchers are unlikely to find a high proportion of all the relevant references. Any effort to increase the number of relevant references retrieved is likely to be at the expense of precision, so that an unacceptably high proportion of irrelevant references may need to be reviewed.

Our systematic review indicates that Medline searching for randomised clinical trials achieves a median precision of about 33%. Sensitivity was highest when precision was at or below 35%, and decreased as precision increased (figure). The point at which to balance precision and sensitivity must be decided by the individuals performing the systematic reviews. For example, some studies included in our review knowingly omitted terms that would have increased sensitivity (COMPARATIVE STUDY, for example) because precision would have been severely compromised. Sensitivity may also vary within subject matter categories and journals searched, although this has not been examined systematically.

From the study of searching for vision trials, we learned that 18% of the relevant Medline articles and 20% overall were not in English. Most (29/37) of the articles that remained unclassified were not reported in English. The proportion of randomised clinical trials overall that are written in languages other than English is no doubt higher than our experience from Medline suggests. Had we searched Embase (the Reed-Elsevier Excerpta Medica database), which includes many European language journals not indexed for Medline, we would undoubtedly have identified additional trials in other languages. These findings imply that excluding from a meta-analysis studies published in languages other than English or limiting a search to Medline will result in more than a trivial number of studies being omitted. A comprehensive search of several databases and non-English publications may lead to considerable translation expenses, depending on the project undertaken. Whether an article reports a randomised clinical trial may not be apparent even after an article has been translated. We recommend not translating the entire article until someone who is able to read the language ascertains that it is or may possibly be a randomised clinical trial.

IMPROVING RETRIEVAL THROUGH MEDLINE

Retrieval of randomised clinical trials through Medline can be improved in three ways:

improve terminology used in reports so that it is clear that they describe the results of a randomised clinical trial;

improve indexing so that all randomised clinical trials are indexed with the appropriate publication type term RANDOMIZED CONTROLLED TRIAL; and

improve strategies used to search for randomised clinical trials. Use of truncation and both subject matter and methods textwords are particularly important.

The first strategy relies on editors taking the lead. Until editors require adequate descriptions of study methodology in the title or abstract, as well as the methods section, of every article published we cannot expect adequate indexing. Indexers can apply a study design term only if authors explicitly describe the design. For example, a tag of RANDOMIZED CONTROLLED TRIAL cannot be applied if authors never state that the study was randomised. Authors and editors must consider the indexing when they are writing, taking special care that the title and abstract are informative and precise; use of structured abstracts should facilitate this.

Improved indexing relies on training and quality measures taken by the National Library of Medicine and also on the availability of indexing terms. Recent changes in Medline indexing should result in more sensitive searches in the future.

Searchers can improve retrieval of randomised clinical trials from Medline in several ways. In our vision study we used a two stage search strategy. We searched one year first, identified the journals that published randomised clinical trials, hand searched those journals for a single year, identified additional MeSH and freetext terms that would have proved useful in identifying articles, and performed a second Medline search incorporating these new terms. Hand searching journals (in our study, 66 journals for 1988) to create a gold standard might seem daunting, but it does provide information on the indexing of the randomised clinical trials that were not picked up by the first search. Modifications of this approach are possible. For example, one could combine the results of a first Medline search with trials identified by other means (for example, review of lists of references incorporated in reports of trials identified by the first search). Jadad and McQuay suggest that combining a Medline search with a selective hand search just of published conference abstracts (such as are found in journal supplements) and letters is a reasonable approach when funds for hand searching are limited.[9] More research is needed as to whether performing a two stage search, such as described here, or a single Medline search plus hand search will achieve a similar sensitivity.

Advances have been made in developing new searching strategies, primarily through comparison of existing strategies and their results. Our review of the strategies used for 12 of the 15 studies included in our meta-analysis found that limited use of textword searching was the most consistent defect. (Our evaluation is potentially biased, however, because we are also authors or advisers on search strategies designed for several of the articles included in the systematic review.)

UNPUBLISHED TRIALS

Improving Medline searching does not address the issue of how best to identify the 25-50% of trials started but never published,[25] or the problem of identifying reports published in non-Medline journals. Linking international databases (Medline and Embase, for example) may help. In addition, access to "grey" literature (literature not formally published, such as research reports, policy documents, dissertations, and conference abstracts) remains difficult. The advent of specific grey literature databases such as SIGLE, produced by the European Association for Grey Literature (a group of national information and documentation centres devoted to providing access to such literature) has improved the situation, but much research remains inaccessible. This literature is not insignificant, and trials reported there may not see full publication elsewhere: on average, only 50% of abstracts reporting the results of randomised clinical trials reach full publication.[24]

Although some authors advocate not making the effort to identify unpublished trials because the data have not undergone peer review,[26] excluding data from unpublished trials will lead to a loss in precision of the estimate of an effect size. In addition, failure to publish is associated with "negative" results (results that are not statistically significant); this association results in a publication bias.[25] Publication bias has serious implications for unbiased data collection for a systematic review. Publication bias extends beyond the failure to publish a report: Stewart and Parmar found that data are selectively omitted from published articles, and they recommend that all systematic reviews be based on data on individual patients rather than published reports.[27] The conditions required for this approach, however, are not available to most reviewers.

Those planning to undertake meta-analyses should not underestimate the difficulty or expense of performing a well conducted systematic review. There is no question that the choice of methods used for data collection is the key to the validity of such a review. Right now, the only alternatives to electronic searching are development of trials registers and use of hand search. Both are costly. But if health care is to be based on all available evidence rather than selected evidence, these costs must be borne or the reviews may be misleading.

REGISTERS OF RANDOMISED CLINICAL TRIALS

Systems for registration of trials should be available to facilitate unbiased data collection for systematic reviews. We are currently participating in the Cochrane Collaboration[28] and are in the process of developing

such a register in collaboration with the National Library of Medicine. Cooperation from investigators and journal editors, as well as support from funding agencies, will be needed to make this effort a success.

The first step is to identify all published randomised clinical trials through electronic and hand searching of the literature. We have devised a generalised search strategy (Appendix), based on the results of this review, that will be used to develop the core of the register. The strategy is in three stages: stage one (sets 1-8) includes terms with high precision, stage two (sets 9-24) includes terms with moderate precision, and stage three (sets 25-34) includes terms with low precision but which provide optimal sensitivity. Each stage is limited to exclude reports solely of animal studies, but retains reports indexed as human and animal, and neither human nor animal.

Journal editors are being asked to arrange for hand searches of their own journals to ensure that all randomised clinical trials published in them will be included in the register and thus will have the best opportunity for inclusion in systematic reviews. The University of Maryland will coordinate the activities involved in identifying reports of randomised controlled trials and ensure that these reports are forwarded to the National Library of Medicine for retagging with an appropriate "publication type" term. In addition to the existing publication type term RANDOMIZED CONTROLLED TRIAL, which was introduced in 1991, the National Library of Medicine has agreed to introduce a new publication type term, CONTROLLED CLINICAL TRIAL, from January 1995. This will be used to tag all reports in Medline that meet the Cochrane Collaboration's criteria for defining a controlled trial but do not meet the library's criteria for indexing under the more specific term RANDOMIZED CONTROLLED TRIAL. Both terms will be applied retrospectively to reports identified by the electronic and handsearching activities described above. Tagging with the new term should become less important as editors insist on explicit descriptions of study methodology which will enable accurate tagging with RANDOMIZED CONTROLLED TRIAL.

The second step of development is prospective registration of all randomised clinical trials. Many trial registers in specific subject areas already exist,[29][30] and the International Collaborative Group on Clinical Trials Registries has been working towards such a goal for several years.[31]

The National Institutes of Health sponsored an international conference in December 1993 entitled "An evidence-based health care system: the case for clinical trial registries," focusing on all aspects of trial registration and bringing together many of the leaders in this field. There is considerable optimism that the scientific, medical, and information communities are moving towards making unbiased data collection for systematic reviews more possible. The implications for evidence based health care are great.

1 Dickersin K, Berlin J. Meta-analysis: state-of-the-science. *Epidemiol Rev* 1992;14:154-76.
2 Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. *Controlled Clin Trials* 1987;8:343-53.
3 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DW. Publication bias in clinical research. *Lancet* 1991;337:867-72.
4 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 1992;263:374-8.
5 Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* [serial online] 1993 Apr 28;Doc No 50:4967 words;53 paragraphs.
6 Scherer R, Dickersin K, Kaplan E, Min TI. Initiation of a registry of randomized clinical trials in vision research. *Invest Ophthalmol Vis Sci* 1992;33:1323.
7 Scherer R, Dickersin K, Kaplan E. The accessible biomedical literature represents a fraction of all studies in a field. In: Weeks RA, Kinser DL (eds). *Editing the refereed scientific journal.* New York: IEEE Press, 1994;24:120-5.
8 Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Comparison of Medline searching with a perinatal trials database. *Controlled Clin Trials* 1985;6:306-17.
9 Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* [serial online] 1993;Doc No 33:3973 words;39 pargraphs.

Appendix

Optimally sensitive Medline search strategy for identifying randomised clinical trials

Set No	Term searched or sets combined
1	RANDOMIZED-CONTROLLED-TRIAL in PT
2	RANDOMIZED-CONTROLLED-TRIALS
3	RANDOM-ALLOCATION
4	DOUBLE-BLIND-METHOD
5	SINGLE-BLIND-METHOD
6	1 or 2 or 3 or 4 or 5
7	TG-ANIMAL not (TG=HUMAN and TG=ANIMAL)
8	6 not 7
9	CLINICAL-TRIAL in PT
10	explode CLINICAL-TRIALS
11	(clin* near trial*) in TI
12	(clin* near trial*) in AB
13	(singl* or doubl* or trebl* or tripl*) near (blind* or mask*)
14	(13 in TI) or (13 in AB)
15	PLACEBOS
16	placebo* in TI
17	placebo* in AB
18	random* in TI
19	random* in AB
20	RESEARCH-DESIGN
21	9 or 10 or 11 or 12 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22	TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
23	21 not 22
24	23 not 8
25	TG=COMPARATIVE-STUDY
26	explode EVALUATION-STUDIES
27	FOLLOW-UP-STUDIES
28	PROSPECTIVE-STUDIES
29	control* or prospectiv* or volunteer*
30	(29 in TI) or (29 in AB)
31	25 or 26 or 27 or 28 or 30
32	TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
33	31 not 32
34	33 not (8 or 24)

Format shown is for SilverPlatter version 3.10. Upper case denotes controlled vocabulary; lower case denotes freetext terms. Those wishing to run this search strategy are recommended to seek the advice of a trained medical librarian.

10 Gøtzsche PC, Lange B. Comparison of search strategies for recalling double-blind trials from Medline. *Dan Med Bull* 1991;38:476-8.
11 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* (in press).
12 Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Controlled Clin Trials* 1985;6:271-9.
13 Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus Medlars searches. *Controlled Clin Trials* 1988;9:23-31.
14 Silagy C. Developing a register of randomised controlled trials in primary care. *BMJ* 1993;306:897-900.
15 Jadad AR, McQuay HJ. Be systematic in your searching. *BMJ* 1993;307:66.
16 Ohlsson A. Treatments of preterm premature rupture of the membranes: a meta-analysis. *Am J Obstet Gynecol* 1989;160:890-906.
17 Daya S. Efficacy of progesterone support for pregnancy in women with recurrent miscarriage. A meta-analysis of controlled trials. *Br J Obstet Gynaecol* 1989;96:275-80.
18 Hofmans EA. The results of a Medline search. The accessibility of research on the effectiveness of acupuncture. II. *Huisarts Wet* 1990;33:103-6.
19 Kirpalani H, Schmidt B, McKibbon KA, Haynes RB, Sinclair JC. Searching Medline for randomized clinical trials involving care of the newborn. *Pediatrics* 1989;83:543-6.
20 Kleijnen J, Knipschild P. The comprehensiveness of Medline and Embase computer searches. Searches for controlled trials of homeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. *Pharmaceutisch Weekblad Scientific Edition* 1992; 14:316-20.
21 Lacy JB, Ohlsson A. The administration of intravenous immunoglobulins for prophylaxis or treatment of infection in preterm infants: meta-analyses. (Submitted for publication, 1993.)
22 Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery* 1994;115:707-12.
23 United States Department of Health and Human Services, Public Health Service, National Institutes of Health. *Fact sheet, technical services division.* Bethesda, MA: National Library of Medicine, 1993:March.
24 Scherer R, Dickersin K, Langenberg P. Full publication of results initially presented as abstracts: a meta-analysis. *JAMA* 1994;272:158-62.
25 Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann NY Acad Sci* 1993;703:135-48.
26 Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam P. Meta-analysis of clinical trials as a scientific discipline. II. Replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987;6:733-44.
27 Stewart L, Parmar M. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418-22.
28 Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992;305:786-8.
29 Easterbrook PJ. Directory of registries of clinical trials. *Stat Med* 1992;11: 345-423.
30 Dickersin K. Why register trials?—revisited. *Controlled Clin Trials* 1992;13: 170-7.
31 Ad Hoc Working Party of the International Collaborative Group on Clinical Trials Registries. Position paper and consensus recommendations on clinical trial registries. *Clinical Trials Meta-analysis* 1993;28:255-66.