

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA**

UMI[®]
800-521-0600



Université d'Ottawa • University of Ottawa

**Assessing the Quality of Reporting
in
Meta-analyses of Randomized Controlled Trials**

by

© Beverley Julia Shea

**Thesis submitted to
The School of Graduate Studies and Research
in partial fulfilment of the requirements for
The M.Sc. Degree in Epidemiology
University of Ottawa**

July 1999



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

395 Wellington Street
Ottawa ON K1A 0N4
Canada

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46609-4

Canada

ABSTRACT

Evidence-Based Health Care is becoming increasingly popular, and in some settings is influencing the direction of health-care policy. If Meta-analyses (MAs) are to continue to be useful, consideration must be given to how they are reported.

This study was carried out to: a) assess the quality of reporting and general characteristics of paper-based MAs published from 1977 to 1995; b) assess over time; c) compare with Cochrane Collaboration (i.e., electronically published) MAs.

Paper-based MAs were randomly selected from a database of systematic reviews. Cochrane Collaboration MAs were randomly selected from the Cochrane Database of Systematic Reviews, Issue 3, 1996. Data were analyzed using univariate and multivariate approaches.

One hundred and fifty-one MAs were reviewed. The overall mean quality score of MAs published in the 'Early Years' was 3.04 (95% Confidence Interval (CI) 2.53,3.56); 'Later Years' 3.35 (95% CI 2.83,3.87), and the Cochrane Library MAs 3.42 (95% CI 2.92,3.93).

This study reveals that there has been a trend, over time, towards improvement in the reporting quality of paper-based published MAs and that the reporting quality of (Cochrane) electronically based MAs are no different from that of paper-based MAs.

ACKNOWLEDGEMENTS

Many people have encouraged me to do post-graduate work, and helped me complete this thesis. I welcome this opportunity to express my appreciation to them.

By her birth, my daughter, Julia Kylie Donahue, strengthened my determination to get my thesis done, so that I could have more time to spend with her. My husband, Brian, has unselfishly put up with my many absences while I was hitting the books. Thank you, Brian, for your ongoing support, and especially for my dictionary!

Without the supportive love, and encouragement of my mother, Viola Shea, I could not have completed this task. Thank you, Mom, for everything! During his lifetime, my father, James Shea, encouraged me to set challenging goals and believe that I could achieve them. My three brothers, Chris, Terry and Eugene, have challenged me to match the pace they have set.

Candyce Hamel, our research assistant, never once refused a request for help in retrieving articles, fixing references and checking tables. She has faithfully picked up on the many things that I missed, and never complained about my constant anguish!

I would also like to thank those who gave me guidance and support along the way. My thesis supervisor's, David Moher and Dr. Peter Tugwell, went so far as to permit me to contact them at any time, day or night. Ba Pham gave ongoing statistical advice, and Dr. Ian Graham put up with my request to fix my mistakes in SpSS, and kindly offered comment on draft versions of my thesis. Alison Jones provided the database of MAs and helped retrieve the articles, Kylie Hugo and Ron and Cynthia Habinski made excellent editorial suggestions, and encouraged me by believing that I would one day complete my task. Dr. Bill Cole encouraged me to take this task on and Drs. Andreas Laupacis and George Wells of the Clinical Epidemiology Unit of the Loeb Research Institute, provided working space and tools, and ongoing encouragement. Bernhard Gibbis helped with translation work, and Eugene Delabays and Luc Berthiaume participated in pilot projects.

A special thanks to Dr. Andy Oxman and Dr. Henry Sacks for their permission to use their assessment tools.

And finally, in memory of my friend Krystyna Malyk, who also helped with the translation of articles.

The Department of Epidemiology and Community Medicine of the University of Ottawa, authorized me to conduct this Project.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	VIII
LIST OF FIGURES	IX
LIST OF TABLES.....	X
LIST OF APPENDICES.....	XII

TEXT

1.	INTRODUCTION	1
1.1	Reporting of meta-analysis	3
1.2	Choosing the quality-assessment instruments	5
1.2.1	Oxman and Guyatt (OQAQ)	7
1.2.2	Sacks	9
1.3	Aims and objectives.....	9
1.3.1	Assess the quality of reporting in paper-based published meta-analyses of the 'early years' and 'later years' (1977-1995)	9
1.3.2	Assess the quality of reporting in paper-based meta-analyses of the 'early years' versus the 'later years'	10
1.3.3	Assess the quality of reporting in paper-based versus electronically published meta-analyses	10
1.3.4	Development of effective search strategies for retrieving meta-analyses	10
1.3.5	Compare the results of this study with results reported in historical published literature.....	10
2.	METHODS	11
2.1	Selecting meta-analyses.....	11
2.1.1	Initial search strategy	11
2.1.2	Second search strategy: Moher et al (1999) database of known meta-analyses.....	12
2.2	Choosing appropriate time periods.....	16
2.2.1	Paper-published meta-analyses from the 'early years'	16
2.2.2	Paper-published meta-analyses from the 'later years'	16
2.2.3	Electronically published meta-analyses from The Cochrane Library	16
2.3	Calculation of sample size	17
2.3.1	For pilot studies	17
2.3.2	For the major research objectives	18
2.4	Generating random lists of meta-analyses	18
2.4.1	Masking.....	18
2.5	Data collection forms.....	18
2.5.1	Translation	19

2.6	Pilot studies.....	19
2.6.1	Pilot One: Inter-rater reliability	19
2.6.2	Pilot Two: Bias check	19
2.7	Analysis.....	19
2.7.1	Frequency data	20
2.7.2	Univariate analysis.....	20
2.7.2.1	Partitioning of degrees of freedom in r x 2 tables.....	20
2.7.3	Multivariate analysis	21
2.7.4	Comparison of results of this study with data of Sacks et al	22
3.	RESULTS	23
3.1	Results of pilot studies.....	23
3.1.1	Pilot One: Inter-rater reliability	23
3.1.2	Pilot Two: Bias Check	24
3.2	Results of Oxman and Guyatt scale (1991) application	24
3.2.1	Paper-based published meta-analyses from the 'early years'	24
3.2.2	Paper-based published meta-analyses from the 'later years'	27
3.2.3	Electronically published meta-analyses from the Cochrane Library	30
3.2.4	Comparing paper-based meta-analyses published in the 'early years' with those published in the 'later years'	33
3.2.5	Comparing paper-based meta-analyses and electronically published meta-analyses	35
3.3	Sacks et al checklist (1987) results	40
3.3.1	Paper-based published meta-analyses from the 'early years'	40
3.3.2	Paper-based published meta-analyses from the 'later years'	43
3.3.3	Electronically published meta-analyses from <u>The Cochrane Library</u>	46
3.3.4	Comparing paper-based published meta-analyses from the 'early years' with those of the 'later years'	49
3.3.5	Comparing 'later years' paper-based and (Cochrane) electronically published meta-analyses	52
3.4	Multivariate analysis (for all statistically significant and substantively significant greater than 10%).....	56
3.4.1	Oxman and Guyatt - year of publication	56
3.4.2	Sacks - year of publication.....	56
3.5	Comparison to Sacks et al (1987) data	58
3.6	Summary of results	60
4.	DISCUSSION AND CONCLUSIONS	65
4.1	Major research objectives	65
4.1.1	Objective one - Quality of reporting on paper-based meta-analyses (1977-1995).....	65
4.1.2	Objective two - Quality of reporting on paper-based meta-analyses from the 'early years' versus the 'later years'	67

4.1.3	Objective three - Quality of reporting on paper-based versus electronically published meta-analyses.....	67
4.2	Minor research objectives.....	68
4.2.1	Objective four – Development of effective search strategies for retrieving meta-analyses	68
4.2.2	Objective five – Comparison of the findings of this study with findings reported in the published literature	68
4.3	Study limitations	69
4.4	Implications of results.....	70
4.4.1	Solutions	72
4.5	Significance of research.....	73
REFERENCES		76
APPENDICES		80

List of Abbreviations

CC	Cochrane Collaboration
CDSR	Cochrane Database of Systematic Reviews
CI	Confidence Interval
CL	The Cochrane Library
CONSORT	Consolidation of Structured Reporting of Randomized Trials
DARE	Database of Abstracts of Reviews of Effectiveness
Early Years	1977-1989
EBHC	Evidence-Based Health Care
ELMA	Electronically published Meta-analyses
EMBASE	Excerpta Medica Database
ICC	Inter Class Correlations
Later Years	1990-1995
MA	Meta-analysis
MARCT	Meta-analysis of Randomized Controlled Trials
MEDLINE	MEDLARS Online
MeSH	Medical Subject Headings
OQAQ	Overview Quality Assessment Questionnaire
OR	Odds Ratio
PBMA	Paper-Based Meta-analyses
QUOROM	Quality of Reporting of Meta-analyses
RCT	Randomized Controlled Trial
S Plus	Statistical Plus Software
SPSS	Statistical Package for Social Sciences
SR	Systematic Review
SS	Search Strategy

List of Figures

Figure 1	Flow chart for Selection of Meta-analyses	12
Figure 2	Flow chart for Database of Meta-analyses.....	15
Figure 3	Average Scores (Overall Q10, 1 to 7) for 'Early' and the 'Later' Years and Electronically Published (Cochrane) Meta-analyses	38
Figure 4	Paper-based Published Versus Electronically Published (Cochrane).....	39

List of Tables

Table 1 Oxman and Guyatt Overview Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1977-1989	26
Table 2 Oxman and Guyatt Overview Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1990 and 1995	29
Table 3 Oxman and Guyatt Overview Quality Assessment Questionnaire Published in the Cochrane Database of Systematic Reviews between 1993-1996	32
Table 4 Oxman and Guyatt Overview Quality Assessment Questionnaire Comparisons of Paper-based Data Published Between 1977 and 1989, and Paper-based Data Published between 1990 and 1995	34
Table 5 Oxman and Guyatt Overview Quality Assessment Questionnaire: Comparison of Paper-based Data Published Between 1990 and 1995, and Cochrane Data Published Between 1993-1996	36
Table 6 Sacks Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1977 and 1989	42
Table 7 Sacks Quality Assessment Questionnaire Published in Peer Review Journals Between 1990 and 1995	45
Table 8 Sacks Quality Assessment Checklist Published in the Cochrane Database of Systematic of Reviews Between 1993-1996	48
Table 9 Sacks Quality Assessment Checklist: Comparisons of Paper-based Data Published Between 1977 and 1989 and Paper-based Data Published Between 1990 and 1995	50
Table 10 Sacks Quality Assessment Checklist: Comparisons of Paper-based Data Between 1990 and 1995 and Cochrane Data Between 1993 and 1996	54
Table 11 Sacks Quality Assessment Checklist: Comparison of Sacks Data Published Between 1983 and 1987 and Data included in this study As Published Between 1977 and 1989	59

Table 12	Oxman and Guyatt Overview Quality Assessment Questionnaire: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results	62
Table 13	Sacks Quality Assessment Checklist: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results.....	63
Table 14	Sacks Quality Assessment Checklist: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results.....	64

List of Appendices

Appendix A	Cochrane Collaboration Glossary	80
Appendix B	List of Checklists and Scales	85
Appendix C	Oxman and Guyatt's Scale (OQAQ)	87
Appendix D	Sacks Quality Assessment Checklist for Meta-Analyses	89
Appendix E	Set Search.....	90
Appendix F	Enhanced Version of Oxman and Guyatt's Scale (OQAQ)	91
Appendix G	Enhanced Version of Sacks Quality Assessment Checklist for Meta-Analysis	94
Appendix H	Guidelines for completing Sack Quality Assessment Form	97
Appendix I	Logistic Regression (Sacks Question 1)	99
Appendix J	List of Meta-Analyses Included in this Study	102

1. INTRODUCTION

People in progressive societies, e.g., around the world want good quality health care, because of its expected impact on their life quality and expectancy. The same motivation has led to strong interest in discovering better ways of meeting health needs. The proliferation of medical opinions, procedures, treatments and drugs, both ancient and modern, requires ordinary people to make decisions as to who to trust for their medical care, and what kind of advice and treatment to accept. Medical academics and practitioners are concerned that their decisions about the medicine that they teach and practice be soundly based. This has led to the widespread use of randomized controlled trials (RCTs) to study the effectiveness of specific treatments, procedures and drugs, or compare their effectiveness to that of standards of known efficacy.

Of course, RCTs are conducted to gain reliable information on which to base medical decisions. When the evidence gathered in this way is strong and its implications clear, it is hoped that it will influence the medical decisions of professionals, and shape the formation of national and international health-care policy. Evidence-based health care (EBHC) is becoming increasingly popular, and in some settings, at least, is influencing the direction of health-care policy.

Unfortunately, the reported results of RCTs are not always clear, convincing or unequivocal. Moreover, the reported results of various trials, compared side by side, have often been confusingly heterogeneous or flatly contradictory. This has given rise to the recognition, within the medical and scientific communities, of the need for the careful, objective and methodical reappraisal of the pooled results of controlled trials, and of the conclusions drawn from those results by the researchers who conducted them. It has come to be accepted that such after-the-fact studies or meta-analyses (MAs) make it possible to extract from pooled studies information of a higher degree of generalization and reliability than the individual trials on which they are based.

Meta-analysis provides a systematic and explicit method for synthesizing evidence; an

overall quantitative estimate (and confidence intervals) derived from the individual studies; and, early evidence as to the effectiveness of treatments, thereby reducing the need for continued study (Moher 1995). MAs offer several potential advantages including: 1) increased statistical power to detect significant differences by increasing sample size, especially for subgroups; 2) improved estimates of effect size and precision; 3) a means of resolving uncertainty when reports disagree; 4) an ability to answer research questions not posed by individual trials; and 5) initiate improvements in the quality of the primary research question (Sacks 1987).

Meta-analyses were used initially to synthesize data about relatively simple health care interventions (e.g., drug interventions) to answer questions such as ‘whether the treatment produced benefit on average’ (Bailey 1987). This approach has proved very useful for examining the effectiveness of clinical interventions with relatively small effect sizes. Increasingly, however, both MAs and Systematic Reviews (SRs) are being undertaken to evaluate more complex health care interventions, the effects of which may be modified by a wide variety of factors. These two terms have been used interchangeably for several years. Using the glossary developed by the Cochrane Collaboration (1995), a systematic review is defined as a review in which evidence on a topic has been systematically identified, appraised and summarized according to predetermined criteria. Such reviews can be systematic (i.e., taking steps to reduce bias) without using statistical synthesis (i.e., MA) to reduce imprecision (Appendix A).

Health care providers and consumers need to be confident that results from MAs are as free as possible from bias (i.e., systematic error). MAs with minimal bias are more valid, and hence should, ideally, more strongly influence health care practice and policy.

One way to assess the merits of an MA is to examine the validity of its report. However, a scientific report may not reflect how its authors conducted their MA as much as it does their ability to write comprehensibly. Although there is sparse data addressing this point, it appears that a scientific report is a reasonable marker for how a project was conducted. Liberati et al. (1986) assessed the quality of 63 reports of randomized trials relating to breast

cancer using a scale with a maximum score of 100 points. Their preliminary findings yielded an average quality score for these reports of 50% (95% CI: 46, 54). The team then interviewed 62 of the corresponding authors of these reports to ascertain whether pertinent information in the manuscripts submitted for journal consideration was removed prior to publication. Even with the additional information obtained from these interviews, the average quality score increased only marginally to 57%.

1.1 Reporting of MA

Over the past decade, the number of MAs published annually has increased 500 fold (Moher 1997). There was a dramatic increase in the number of published MAs between 1955 and June 1992 when a total of 678 were published. A MEDLINE search between July 1992 and 1995 yielded 527 articles with MA in the title and 976 with MA as a key word (Moher 1995).

Little is known about the quality of reporting in MAs, or even why it is important. By comparison, a plethora of articles has been published on the quality of reporting in RCTs. A review of eighty-six English language MAs by Sacks et al. (1987) assessed every report on fourteen items from six content areas believed to be critical in the conduct and reporting of MA. These items included study design, combinability, control of bias, statistical analysis, sensitivity analysis and problems of applicability. They found that only 24 of the 86 (28%) MAs addressed all six contents areas. This survey was updated in 1992 with little change in the results obtained (Sacks 1996).

Silagy (1993) surveyed 28 systematic reviews published in primary care journals during 1991 using eight methodological criteria understood to be important in the reporting of systematic reviews (SRs). Each criterion had a maximum score of two for a total score of 16. Silagy reported that only 25% of the articles obtained a total score higher than eight.

In a recent article, Jadad and McQuay (1996) reviewed 80 MAs using a validated assessment tool (Oxman & Guyatt, 1991), which included a question assessing the overall scientific quality of each MA. The scoring range for this question was from one to seven with higher

scores indicating superior quality. The authors reported a median score of 4. Studies with an Oxman and Guyatt Scale score between three and four are considered to have major flaws. The authors also found that MAs with positive results were more likely to have lower scores. This would appear to suggest a worrisome link between “good” results and poor quality. Assendelft et al. (1995) used a similar approach to assess the quality of 51 SRs of spinal manipulation. They too reported a low median score (23 out of a maximum score of 100). However, they reported that reviews with statistically “positive” conclusions tended to have higher quality scores, a finding that runs counter to the conclusions of Jadad and McQuay (1996).

This evidence suggests that unless more attention is devoted to improving the reporting of MAs, we are likely to encounter situations as lamentable as those occasionally seen in the reporting of RCTs. It is urgent that this problem be addressed. The ultimate test of an MA is whether its report inspires confidence in the reader that it accurately reflects the process followed during its various stages. The report itself is critical, since it is the information source most relied upon by clinicians in their management of patients (Moher 1997).

Improving the quality of reporting in MAs is also very important to health policy makers and members of the Cochrane Collaboration. The Cochrane Collaboration is an international multi-disciplinary collaboration established in 1993. Its avowed task is to prepare, maintain and disseminate systematic, up-to-date reviews of randomized controlled trials of health care, and, when RCTs are not available, to review the most reliable evidence from other sources.

If MAs are to continue to be useful, serious consideration must be given to how they are conducted and reported. As numerous researchers around the world develop guidelines for the conduct of MAs in all areas of health care, it is important to examine how these summaries are produced. It is imperative to determine not only the sources of heterogeneity being introduced within this pooled data, but also the differences in how meta-analysts conduct their reviews.

Recent studies indicate widespread linguistic limitations or bias in the work of many meta-

analysts. The majority of MAs were found to be based exclusively on English language RCTs. It has been shown that 78% have language restrictions in their use of sources (Gregoire 1995). The majority (93%) of completed MAs exclude trials reported in languages other than English. Yet some empirical evidence indicates that there is no difference in the quality of reporting in English and non-English language RCTs (Moher 1996). Egger et al. have reported similar results (Egger 1997).

There are several pressing issues to examine when assessing the quality of reporting in MAs. Traditionally, when a research article, such as a meta-analysis, is submitted for publication there is a formal and consistent process of review. One to three editors are generally employed by the journal to ensure the high quality of published reports. Journals also have a list of experts to whom submitted manuscripts are distributed for review and feedback.

The Cochrane Collaboration has a different peer review process for reviewing and editing submitted material. The Collaboration is organised in what are known as entities. One of these entities is comprised of a number of review groups. Individual review groups have been established to cover specific content areas, and consist of teams of four to 13 editors. One of these takes a leading role as the co-ordinating editor. Under this approach, the editorial teams of individual review groups are ultimately responsible for what is published in the Cochrane Database of Systematic Reviews (CDSR). This is one of the databases included in the electronic journal subtitled "The Cochrane Library". In order for a submission to become a Cochrane review, it must first be accepted by the editorial team responsible for the content area concerned, and then be submitted to the journal by the team via the parent database.

1.2 Choosing the Quality-assessment Instruments

A systematic review of published scales and checklists was performed to take an inventory of available tools. For scales, each item is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to measure across a continuum. A checklist provides an estimate of the overall quality of a systematic review, by using itemized criteria to

assess individual reviews and facilitate their qualitative comparison to one another. Checklist items do not have numerical scores attached to them.

The Literature Search was conducted using Medline from January 1966 to February 1999, by 3 independent searches using the following keywords: meta-analysis, review literature, systematic or quantitative or methodologic review, overview, review, information synthesis, integrative research review, guideline, checklist, tool, scoring, scale, clinimetric, quality, critical reading, methodology and PubMed using the “related articles” function to find others.

The identification and selection of the assessment tools included an initial screening to identify relevance. Potentially relevant articles were reviewed independently by each author and all articles were eligible regardless of language. The article had to be a scale or checklist designed to assess quality of systemic reviews and meta-analyses. All checklists and scales were assessed for: 1) number of items included in tool, 2) type of quality assessed, 3) whether or not article included explicit statement regarding purpose of tool, and 4) time to completion of tool. The data extraction was completed in a group and consensus reached.

The search identified 318 potentially relevant articles. After eliminating duplicates or previously published instruments and those that were not scales or checklists, twenty-six instruments were included in our review; 23 checklists and three scales [Appendix B]. All of the instruments, except one scale (Oxman 1991), have been published. The instruments were developed between 1984 and 1997, indicating a fairly recent interest in this area. The number of items in each instrument ranged from 5 to 101, with only two checklists having more than 35 items. Three checklists and one scale were developed to assess quality assessment within specific domains, such as diagnostic tests. The remaining 20 checklists and two scales were developed to be used with all types of meta-analysis. Fifteen checklists and one scale included an explicit statement regarding the purpose of their instrument. The average time required to assess the quality of a meta-analysis using the checklists and scales was 13 minutes [range: 5 to 30] and 12 minutes

[range: 5 to 20] respectively.

1.2.1 Oxman and Guyatt (1991) (OQAQ)

The Oxman and Guyatt Scale (1991) is called the Overview Quality Assessment Questionnaire (OQAQ). The specified purpose of this instrument is to evaluate the scientific quality (i.e., adherence to scientific principles) of MAs published in medical literature. It is not intended, therefore, to measure literary quality, importance, relevance, originality or other such aesthetic or philosophical attributes of reviews.

The scale was designed for assessing MAs of primary (“original”) research on pragmatic questions related to causation, diagnosis, prognosis, therapy or prevention. The same principles that apply to epidemiological surveys also apply to MAs: a question must be clearly specified; a target population identified and assessed; appropriate information obtained from that population in an unbiased fashion; and conclusions derived, sometimes with the help of formal statistical analysis.

The validity of the OQAQ scale was tested thoroughly at the time of its development. Nine judges used it to assess thirty-six published review articles. These reviews were drawn from three sampling frames: articles highly rated by criteria external to the study, MAs, and a broad spectrum of medical journals. The sensibility (i.e., face validity) of the OQAQ was assessed using a thirteen-item questionnaire. In addition, its sensibility was assessed by 15 randomly selected individuals. Seven a priori hypotheses were also used to assess its construct validity.

Authors’ reports of their methods in personally assessing the thirty-six articles related closely to ratings yielded by corresponding OQAQ items. The mean score yielded by the scale was significantly higher for articles for which the authors’ responses indicated that they had used more rigorous methods. For ten of the thirteen questions used to assess sensibility, the mean rating was five or greater, indicating general satisfaction with the instrument. Six of the seven hypotheses used to test construct validity held true. The OQAQ was therefore

declared a valid measure of the quality of research overviews (Oxman 1991).

When this instrument is applied to an MA, answers to its questions are to be based, as much as possible, on information provided in the article. If the methods employed were reported incompletely relative to a specific item, this item is scored as 'partially'. If no information is provided regarding what was done relative to a particular question, the item is scored as 'can't tell', unless there is information in the article to suggest whether the criterion was met, in which case the item is scored 'yes' or 'no', as appropriate.

The scale is divided into nine areas. Its questions extract information about the following issues: 1) search methods used; 2) comprehensibility of this search strategy; 3) criteria used for inclusion/exclusion; 4) avoidance of bias in the selection of studies; 5) criteria (methodological quality) used for assessing the validity of the included studies; 6) appropriate criteria for assessing validity of all the studies referred to in the text; 7) methods used to combine findings of relevant studies; 8) findings of the relevant studies combined appropriately relative to the primary question; and 9) conclusions drawn by the author(s) supported by the data. A detailed description of the scale and its scoring is provided in Appendix B.

Question 10, the final question on the scale, requires each assessor to rate the overall scientific quality of each report. Possible scores range from one to seven. The score on this question is based on the answers given to the first nine questions. The following guide is recommended by the developers: If the "can't tell" option was used one or more times on these questions, a review was likely to have major flaws (i.e., a score of 4 or lower). If the "no" option was used on questions 2, 4, 6 or 8, the review was likely to have major flaws (i.e., a score of 3 or less, depending on the number and degree of the flaws) (Oxman, 1994).

This instrument was selected based upon its strong face validity and the availability of a published assessment of its construct validity (Oxman, 1991). (See Appendix C)

1.2.2 Sacks

The Sacks checklist was chosen for three reasons: 1) to conduct research similar to that carried out in this area by Sacks et al. (1987); 2) to take this research one step further in a plan to compare data collected over time; and 3) to assess data which is not evaluated by the Oxman and Guyatt Scale (1991).

The Sacks Criteria for assessing quality measures six key areas with a total of twenty-three questions. Appendix D defines each of the key terms used in these questions. The six key areas assessed by this instrument are:

1. Study design (including protocol, literature search, list of trials analyzed and log of rejected trials, treatment assignment, and range of patient characteristics, diagnosis and treatments);
2. Combinability (i.e., of criteria and measurements);
3. Control and measurement of potential bias (i.e., selection bias, data-extraction bias, and sources of support);
4. Statistical Analysis (i.e., statistical methods, statistical errors, confidence intervals, subgroup analyses);
5. Sensitivity Analysis (quality assessment, varying methods, publication bias);
6. Application of results (i.e., caveats, economic impact). Language of included MAs was added to this subset as an additional item.

Sacks et al. (1987) used these criteria to conduct an assessment of the quality of reporting of 86 MAs. Their report was later updated to include an additional 78 MAs (1996). The results of these studies have been summarized previously (Section 1.1).

1.3 Aims and Objectives

Primary research objectives of this study:

1.3.1 Assess the quality of reporting in paper-based published MAs of the 'Early Years' and 'Later Years' (1977-1995).

1.3.2 Assess the quality of reporting in paper-based MAs of the 'Early Years' versus the 'Later Years'

1.3.3 Assess the quality of reporting in paper-based versus electronically published MAs

Secondary research objectives:

1.3.4 Develop effective search strategies for retrieving MAs

1.3.5 Compare the results of this study with results reported in historical published literature

2. METHODS

2.1 *Selecting Meta-analysis*

Two different strategies were employed in retrieving the MAs used for this project. The first selection process involved developing a highly specific search strategy for retrieving MAs, running the search, and reviewing the abstracts retrieved. When this process proved ineffective, a second approach was developed. This second approach involved retrieving MAs from a known database of MAs. A full description will now be given of the development of these two selection processes and the reasons for abandoning the approach initially employed.

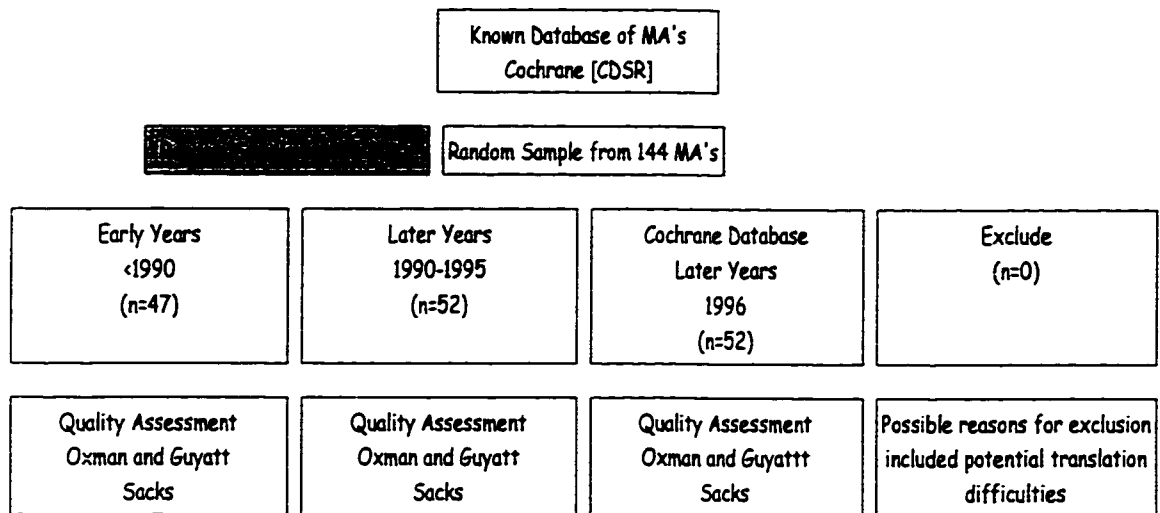
2.1.1 *Initial search strategy*

An attempt was made to develop a search strategy with high sensitivity and specificity in order to select MAs from MEDLINE, an electronic database maintained by the U.S. National Library of Medicine.

This search strategy was developed in collaboration with a library scientist (Jessie McGowan) and other researchers in the area (Alex Jadad, Ann McKibbin and Brian Haynes). The initial strategy included abstracts in all languages for the years 1966 to 1995, and retrieved 5009 abstracts of possible MAs and SRs. Each abstract was numbered and reviewed in detail, to determine which were true MAs (i.e., met the inclusion and exclusion criteria set prior to searching the database). For an MA to be eligible for inclusion in this study, its methods section had to state the type of primary studies included in it. The abstracts did not clearly indicate which reports met this study's definition of meta-analysis. I concluded that the only way to determine the suitability of reports for inclusion in this study would be to obtain the full text of the entire set of articles. Given the limited resources available, this was impractical. When it was learned that a database of known MAs was being developed concurrently by Moher and colleagues, it was decided to retrieve MAs for this study from that database, once it was finalized.

Figure 1

Flow Chart for Selection of Meta-analyses



2.1.2 Second search strategy: Moher et al (1999) Database of known MAs

The following is a description of the methodology used to compile the Moher et al. Database of MAs. (Appendix E)

Phase one: It was decided that, to be considered an MA, an article would have to state: 1) the name(s) of the database(s) searched; 2) the years searched; and 3) the search terms included. The majority of the articles examined failed to report this information. Consequently, it was decided to focus on identifying reports of meta-analysis of RCTs (i.e., MARCTs).

Phase two: A MEDLINE search, using the Search Engine OVID, was carried out from January 1, 1966 to December 31, 1995 to identify MAs (Appendix C). This search included such search terms as Medical Subject Headings (MeSH), text words and publication types. Abstracts retrieved by the search were reviewed. Excluding the studies that were, in fact, SRs was not possible, because the methodology followed was insufficiently described in the

abstracts examined. Unfortunately, the citations were not indexed as “meta-analyses” by the U.S. National Library of Medicine. As a result, it was decided to obtain and read all retrieved articles. As an initial step, hard copies of 50 reviews were randomly reviewed to confirm fulfilment of eligibility criteria.

Phase three: To identify MAs, an electronic search of MEDLINE (OVID) was conducted using the time period of January 1, 1966 to December 31, 1995 (Appendix C). This search strategy included 21 search terms as MeSH, text words and publication types. The MEDLINE search was translated using the appropriate terms to search EMBASE from January 1, 1980 to November 30, 1995. Twenty-five search terms were included. Both search strategies aimed at identifying MAs and SRs published in any language.

Phase four: The Cochrane Database of Systematic Reviews (CDSR), where all Cochrane reviews are published, 1995, Issue 2, was also searched for possible MAs, as was the Database of Abstracts of Reviews of Effectiveness (DARE). Both of these databases are held within The Cochrane Library. DARE, which includes non-Cochrane MAs, did not provide complete bibliographic information for each reference, so hard copies of the papers identified by it could not be retrieved. Consequently, it was decided not to use DARE data in the search for MAs. Current versions of DARE include appropriate sources to facilitate the retrieval of hard copies.

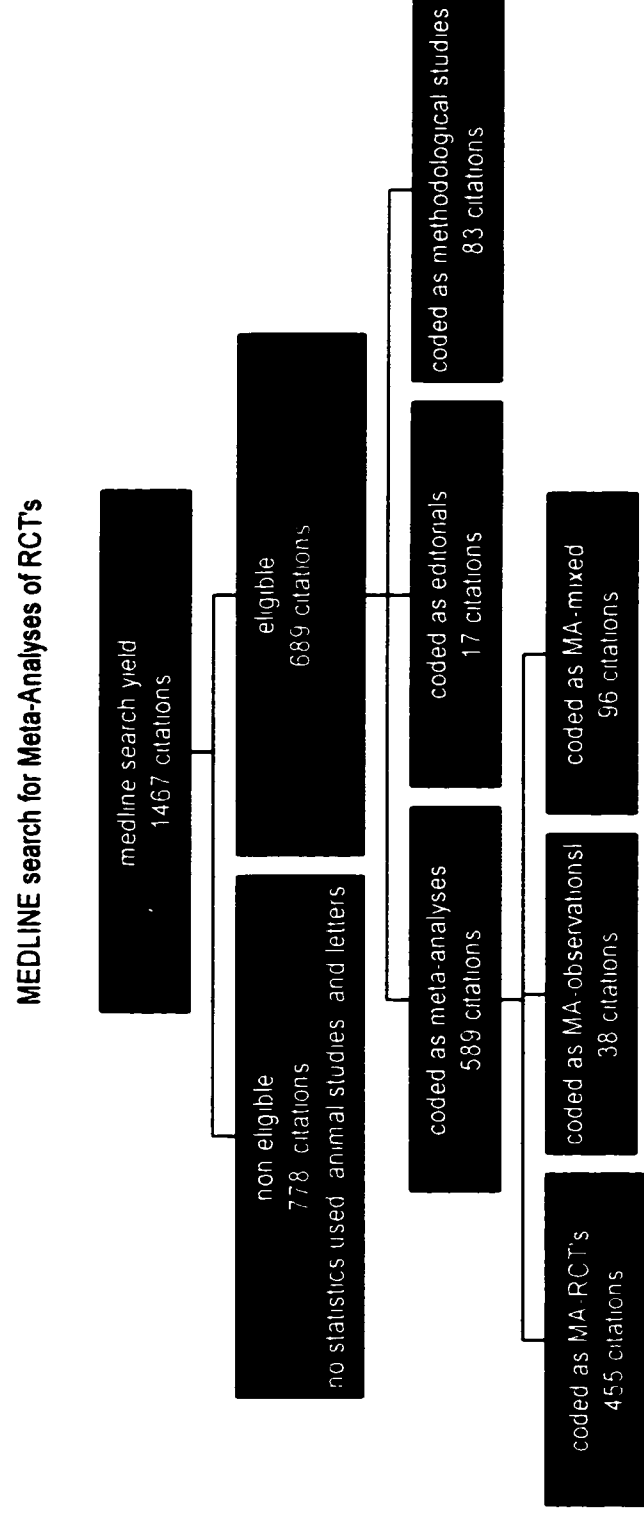
Phase five: Sensitivity and precision were calculated for each MEDLINE search strategy tried. For this purpose, ‘sensitivity’ was understood as the number of true MAs identified by a search method expressed as a percentage of the total number of relevant articles selected by it. ‘Precision’ was understood as the number of MAs identified within a given collection of studies by the MEDLINE search strategy, expressed as a percentage of the total number of MAs known to be included in that collection of studies. Precision was calculated by comparing citations identified by each search strategy with established bibliographic lists of MAs in the American College of Physician’s Journal Club. Based on the results of these quality control efforts, the search strategy was modified and refined to maximise its sensitivity and precision in the selection of titles.

Phase six: Each article identified by the MEDLINE, EMBASE and CDSR searches was evaluated for inclusion based on four criteria: 1) eligibility (Did the article refer to MAs?); 2) publication type (Was the paper an MA, an editorial, or a methodological paper?); 3) primary studies (Did the MA include RCT, observational, or mixed studies?); and 4) type of research question (Was the article focused on treatment, diagnosis, prevention, aetiology, association, prognosis and/or economics?).

Phase seven: The following information was abstracted from each database: number of articles identified by the search strategy; year of publication; number of journals cited; total number of articles coded as MAs; number of articles coded as MARCT; and, observational studies.

Figure 2

Flow Chart for Database of Meta-analyses



2.2 *Choosing Appropriate Time Periods*

2.2.1 *Paper-published MAs from the 'Early Years'*

To measure improvement over time it was important to analyze a data set that included MAs from two different time periods. The 'Early' and 'Later' time periods chosen made it possible to compare results with other studies conducted and validated under similar circumstances. The 'Early Years' represented a sample of MAs published prior to 1989. The earliest MA randomly chosen in this group was from the year 1977. A wider time span was assigned to the 'Early Years' category than to the 'Later Years' category, in order to obtain an adequate sample size, since fewer MAs were published prior to 1989.

2.2.2 *Paper-published MAs from the 'Later Years'*

The data set of MAs from the 'Later Years' included a random set of MAs from 1990 to 1995.

2.2.3 *Electronically published MAs from The Cochrane Library*

Since a Cochrane review goes through an editorial process somewhat different from that of a traditional peer reviewed journal, it was decided to examine potential differences between these two types of reviews. A comparison of the quality of reporting and general characteristics of these two types of reviews was conducted using the scale developed by Oxman and Guyatt (1991) and the checklist developed by Sacks et al. (1987).

The data set derived from the application of these two tools to the traditional (paper-published) MAs provided data to compare with and confirm previous work done in the same area. A similar time period was specified for the selection of electronically published studies. The data set derived from the study of these MAs would provide new scientific knowledge regarding the reporting of Cochrane MA.

Reviews from the CDSR were randomly selected from the 1996 issue since this was the first edition made available on CD-ROM. The reviews in this edition included those completed prior to 1996.

2.3 *Calculation of Sample Size*

2.3.1 *For pilot studies*

Estimating reliability is a common feature of scientific experimentation, since all measurement is subject to error, particularly human-made error. Shrout and Fleiss (1979) concluded that measurement error can seriously affect statistical analysis and interpretation, and so it is important to estimate the amount of such error by calculation of a reliability coefficient (Shrout and Fleiss). Donner and Eliasziw (1987) provided an exact power contour to guide the planning of reliability studies. This was determined where the parameter of interest was the coefficient of interclass correlation ρ derived from a one-way analysis of variance model.

Landis and Koch (1977) have characterized values of reliability coefficients as follows: slight (0-0.20); fair (0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and, almost perfect (0.81-1.00). The contours displayed the required numbers of subjects, K , the number of repeated measurements and n which provided 80% power for testing $H_0: \rho \leq \rho_0$ versus $H_1: \rho > \rho_0$ at the 5% level of significance for selected values of ρ_0 . Since this study required a substantial level of reliability, the equation $H_0: \rho \leq 0.60$ versus $H_1: \rho > 0.60$ was employed, and using the graphs it was estimated that a sample size of nine MAs was required to ensure a sufficient measure of reliability for the pilot study.

Each of the items from both the Sacks checklist (1987) and the Oxman and Guyatt scale (1991) was reviewed and discussed by three raters with regard to its interpretation. Where there were discrepancies, consensus was reached. A priori, it was decided that an ICC of >0.66 would be acceptable. In order to reach agreement on the application of the tools, the inter-rater reliability study was carried out on three separate occasions.

2.3.2 For the major research objectives

When determining an adequate sample size of MAs for this project, consideration was given to its research objectives. An adequate sample size was difficult to calculate, given the absence of data concerning the size of differences to be expected from this study. Therefore, a convenience sample of MAs was taken from the 'Later Years', 'Early Years', and CDSR. The power of this sample was calculated post-hoc.

2.4 Generating Random Lists of MAs

Once the MAs from the known database were retrieved, a random list of numbers was generated using the S Plus software. Separate lists were run for:

- a) each classification category, (i.e., grouped from 1977-1989 and 1990-1995);
- b) all MAs contained in the CDSR in Issue 3, 1996.

2.4.1 Masking

After the lists were generated, full copies of all MAs were retrieved, copied and masked using a 'china marker', to conceal author, institution, and journal. Several published empirical studies address this issue (Chalmers 1983; Moher 1996). Studies have shown that assessments under masked conditions were more likely to yield lower and more consistent scores than assessments made under open conditions (Jadad 1996). The replication of these results in another study (Berlin 1997), which also found that lower scores were assigned under masked conditions, gives strong support to the expectation that bias would be introduced in assessments made under open conditions.

2.5 Data Collection Forms

Data collection forms for the extraction of data included in individual reviews were developed using the scale and checklist selected for use in this study (i.e., Oxman & Guyatt, 1991 & Sacks 1987). Additional items were added to the data collection forms for this study. These added items included: language, year of publication, type of publication, type of MA, number of included subjects, and number of included studies (Appendix F, G and H).

Data forms were also developed including items to collect the information required to analyse:

- a) the quality of reports over time;
- b) the quality of traditional peer-reviewed MAs compared to Cochrane reviews.

2.5.1 Translation

MAs in any language were included in the random selection of MAs from the database. Those in languages other than English (i.e., French, German and Portuguese) were translated into English with the assistance of colleagues.

2.6 Pilot Studies

2.6.1 Pilot one: Inter-rater reliability

The writer and two experts in the field conducted an inter-rater reliability study, to ensure that the checklists were consistently applied. The sample size needed for the inter-rater reliability study was calculated using the methods given by Donner et al. (1987) for addressing sample size requirements for reliability studies. Once reliability was confirmed, the remainder of selected MAs were assessed by individual raters.

2.6.2 Pilot two: Bias check

Due to differences in the physical appearance of the various MAs, it was very difficult to mask whether they were paper or electronically based (i.e., Cochrane). Consequently, a second pilot study, which sampled approximately 20% of the selected MAs, was conducted with the help of a medical student. Its purpose was to determine whether the visible difference between paper and electronically published MAs would lead to bias in the assessment of their quality. This pilot provided a statistical basis for confidence that the results it yielded were not influenced by bias.

2.7 Analysis

All data were entered and verified using Microsoft Excel. Statistical analyses and graphs

displaying the results obtained were produced using SPSS version 8.0 for Windows.

2.7.1 Frequency data

To analyse the quality of reporting and general characteristics of traditional paper-published MAs from 1977-1995, descriptive statistics, including confidence intervals, were used (e.g., to calculate frequency of reporting publication bias). The same process was repeated for the selected Cochrane Collaboration MAs.

2.7.2 Univariate analysis

To calculate the quality of reporting and general characteristics of MAs over time, a univariate analysis was conducted. It compared the quality and general characteristics of MAs published from 1977-1989 with those from 1990-1995. Inferential statistics were used to analyse the data (i.e., a student t-test and a chi-square test). To compare the quality and general characteristics of traditional paper-published MAs with Cochrane MAs, the data were analysed using a chi-square test.

2.7.2.1 Partitioning the degrees of freedom in $r \times 2$ tables

When the values obtained were reviewed, apparently significant differences were observed. To identify which variables contributed to this significance a procedure was performed to partition out the degrees of freedom, so as to separately analyze those variables specifically contributing to the significance. Any contingency table may be partitioned into as many 2×2 sub-tables as there are degrees of freedom in the original table (Siegel 1988). Each of the 3×2 tables was then broken down into 2×2 tables resulting in the 'Yes's' being combined with the 'Partials' versus the 'No's' and the 'No's' being combined with the 'Partials' and compared with the 'Yes's'.

If H_0 is rejected when analyzing an $r \times k$ contingency table, it may safely be concluded that the k groups differ on the measured (row) variable (Siegel 1988). However, although it can be concluded that the k groups are different, the results of the chi-square test by itself do not indicate what the differences are. That is, a significant chi-square test result suggests only

that somewhere in the table the observed frequencies are not simply chance deviations from the expected frequencies. It is important to know precisely where in a contingency table the significance discrepancies lie. The partitioning procedure enables the researcher to discover these differences.

To partition the contingency tables, a series of 2 x 2 sub-tables were constructed. The software followed is presented below:

- 1) The frequencies were constructed into an r x k contingency table, using the k columns for the groups or samples.
- 2) A chi-square was calculated for the 1st partition using the following equation:

$$\chi^2 = N \left(\frac{n_{11}n_{22} - n_{12}n_{21}}{n_{1.}n_{2.}n_{.1}n_{.2}} \right)^2$$

$$C_1 C_2 R_{1+1} \left(\sum_{i=1}^I R_i \right) \left(\sum_{i=1}^{I+1} R_i \right)$$

2.7.3 Multivariate analysis

To examine improvement of quality over time, a multiple logistic regression analysis was conducted using all the statistically and substantively significant values from the univariate analysis. Values exhibiting a difference greater than 10 percentage points were accepted as substantively significant. Following discussions with two experts in the field, this percentage had been designated as the minimum substantive level of difference likely to reflect important differences. For those questions where it was not appropriate to calculate frequencies, a comparison of the means was conducted and reported along with a 95% CI for the difference of the means.

The multiple-logistic-regression model was used to evaluate the significance of the data yielded by the binary comparison of outcomes for the three different categories of MAs. In the model, interaction terms were added to address the following two questions:

1. On average, do time and/or type of publication yield different odds ratios (ORs) relative to the referent category?
2. Do the number of studies and/or the number of patients included in an MA yield different

ORs relative to the referent group?

The statistical model was adjusted to accommodate the two publication types and the two time periods, and determine whether reported effects differed as a function of time or type of publication. The numbers of studies and patients reported on in each MA were included in the model to investigate their effects on the quality of MAs. An odds ratio of less than 1.0 for an interaction term indicated inferior performance on a quality assessment item, and an odds ratio greater than 1.0 indicated a superior performance on a quality assessment item.

2.7.4 Comparison of results of this study with data of Sacks et al.

One objective of this study was to replicate the study by Sacks et al. (1987). This provided an excellent opportunity to do a univariate statistical comparison of the data yielded by the two studies.

3. RESULTS

One hundred and fifty one (151) MAs were reviewed. There were 99 paper-based with 47 classified as 'Early Years' and 52 as 'Later Years'. The rest, 52, were Cochrane Collaboration MAs. The average number of studies examined in each MA was nine (interquartile range 4,18); and the average number of patients included in the composite sample studied by each MA was 754 (interquartile range 222, 2580).

3.1 Results of Pilot Studies

3.1.1 Pilot one: Inter-rater reliability

The writer and two experts in the field conducted a pilot study to test the proposed data extraction forms and quality assessment tools. Agreement among those assessing an MA, is believed to be a very important aspect of MA, since the primary source of unreliability stems from differences in assessing and judging study characteristics (Glass 1981). As recommended by Glass et al. (1981) agreement among assessors was maximised through careful training. All inconsistencies identified were reviewed and resolved at weekly meetings, and assessments were revised after consensus was reached. The pilot study was conducted three times to ensure that a high level of agreement was reached among the raters. An interclass Correlation (ICC) of 0.60 was deemed to be the lowest acceptable level of inter-rater reliability. An ICC was calculated for each question in both the Oxman and Guyatt Scale (1991) and the Sacks Checklist (1987). ICC results greater than 0.60 were obtained for the majority of questions on the Sacks Checklist. There was less agreement on the questions of the Oxman and Guyatt OQAQ. Therefore, each item was reviewed and discussed in detail, with agreement being reached on its scoring in cases where there had been disagreement. This pilot was repeated until an ICC of greater than 0.66 was reached. A final interclass correlation coefficient of 0.88 was obtained for question ten of the MA of the Oxman and Guyatt scale, which addresses the overall quality of an MA. The minor refinements incorporated into the interpretation of the scale (Appendix F) and the checklist (Appendix G and H) permitted a greater degree of precision when performing assessments.

3.1.2 Pilot two: Bias check

Using the Oxman and Guyatt (1991) Scale, the following results were recorded:

1. When reviewing paper-published MAs of the 'Later Years', the medical student found a mean of 3.5 (95% CI 1.8 to 5.2). By comparison, this study reported a mean of 3.4 (95% CI 1.5, 5.3).
2. When reviewing (Cochrane) electronically published MAs, the medical student found a mean of 3.9 (95% CI 2.45, 5.35). By comparison, this study reported a mean of 3.9 (95% CI 2.49 5.31).

3.2 Results of Oxman and Guyatt Scale (1991) Application

3.2.1 Paper-based published MAs from 'Early Years'

MAs from the 'Early Years' contained sufficient data to permit answering "yes" ranging from 19.1% to 93.6% to the various questions on the Oxman and Guyatt Scale.

1. Search methods

The search methods used to find evidence were reported in 28 (59.6%) of the MAs.

2. Comprehensive search

The search strategy used was comprehensive in 16 (34%) of the MAs.

3. Criteria for inclusion/exclusion

Criteria used for deciding which studies to include in the MA were reported in 27 (57.4%) of the MAs.

4. Bias in the selection of studies

The avoidance of bias in the selection of studies was reported in nine (19.1%) of the MAs.

5. Methodological quality

The criteria used for assessing the validity of the included studies were reported in 15 (31.9%) of the MAs.

6. Validity assessed appropriately

Appropriate criteria for assessing the validity of studies were mentioned in 40 (85.1%) of the MAs.

7. Methods used to combine findings

The methods used to combine the findings of included studies were reported in 37 (78.7%) of the MAs.

8. Findings combined appropriately

The findings of the relevant studies were combined appropriately relative to the primary question in 44 (93.6%) of the MAs.

9. Conclusions supported by data

The conclusions reached by authors were supported by the data they cited in 40 (85.1%) of the MAs.

10. Overall quality

The overall quality of MAs from the 'Early Years' received a mean score of 3.04 (95% CI 2.53, 3.56). (See Table 1)

Table 1

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1977 and 1989

<i>Questions</i>	Yes n [%]	Partially or can't tell n [%]	No n [%]
1. Were the search methods used to find evidence reported?	28 [59.6]	10 [21.3]	9 [19.1]
2. Was the search strategy for evidence reasonable comprehensive	16 [34.0]	8 [17]	23 [48.9]
3. Were the criteria used for deciding which studies to include in the overview reported	27 [57.4]	13 [27.7]	7 [14.9]
4. Was bias in the selection of studies avoided	9 [19.1]	20 [42.6]	18 [38.3]
5. Were the criteria used for assessing the validity of the included studies reported	15 [31.9]	4 [8.5]	28 [59.6]
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	40 [85.1]	1 [2.1]	6 [12.8]
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	37 [78.7]	8 [17.0]	2 [4.3]
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	44 [93.6]	3 [6.4]	0
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	40 [85.1]	7 [14.9]	0
10. How would you rate the scientific quality of this overview?	3.04 [95 % CI 2.53,3.56]		

*See Appendix C for a full description of each item

3.2.2. Paper-based published MAs from the 'Later Years'

MAs from the 'Later Years' (1990 – 1995) contained sufficient data to permit answering 'yes' from 21.2 to 90.4% to the various questions on the Oxman and Guyatt Scale.

1. Search methods

The search methods used to find evidence were reported in 35 (67.3%) of the MAs.

2. Comprehensive search

The search strategy used was reasonably comprehensive in 22 (42.3%) of the MAs.

3. Criteria for inclusion/exclusion

The criteria used to decide which studies to include in the MA were reported in 38 (73.1%) of the MAs.

4. Bias in the selection of studies

The avoidance of bias in the selection of studies was reported in 11 (21.2%) of the MAs.

5. Methodological quality

The criteria used for assessing the validity of the included studies were reported in 25 (48.1%) of the MAs.

6. Validity assessed appropriately

Appropriate criteria for assessing the validity of the included studies were reported in 47 (90.4%) of the MAs.

7. Methods used to combine findings

The methods used to combine the findings of included studies were reported in 45 (86.5%) of the MAs.

8. Findings combined appropriately

The findings of the relevant studies were combined appropriately relative to the primary question in 52 (90.4%) of the MAs.

9. Conclusions supported by data

The conclusions the authors reached were supported by the data they adduced in 45 (86.5%) of the MAs.

10. Overall quality

The overall quality of MAs from the 'Later Years' received a mean score of 3.35 (95% CI 2.83, 3.87 – see Table 2).

Table 2

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1990 and 1995

Questions	Yes n [%]	Partially or can't tell n [%]	No n [%]
1. Were the search methods used to find evidence reported?	35 [67.3]	6 [11.5]	11 [21.2]
2. Was the search strategy for evidence reasonable comprehensive	22 [42.3]	11 [21.2]	19 [36.5]
3. Were the criteria used for deciding which studies to include in the overview reported	38 [73.1]	7 [13.5]	7 [13.5]
4. Was bias in the selection of studies avoided	11 [21.2]	27 [51.9]	14 [26.9]
5. Were the criteria used for assessing the validity of the included studies reported	25 [48.1]	5 [9.6]	22 [42.3]
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	47 [90.4]	2 [3.8]	3 [5.8]
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	45 [86.5]	4 [7.7]	3 [5.8]
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	47 [90.4]	5 [9.6]	0
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	45 [86.5]	6 [11.5]	1 [1.9]
10. How would you rate the scientific quality of this overview?	3.35 [95% CI 2.83,3.87]		

*See Appendix C for a full description of each item

3.2.3 *Electronically published MAs from The Cochrane Library*

MAs from The Cochrane Library (1996) contained sufficient data to permit answering 'yes' from 26.9 to 100% to the various questions on the Oxman and Guyatt Scale.

1. *Search methods*

The search methods used to find evidence were reported in 18 (34.6%) of the MAs.

2. *Comprehensive search*

The search strategy was reasonably comprehensive in 14 (26.9%) of the MAs.

3. *Criteria for inclusion/exclusion*

The criteria used for deciding which studies to include in the MA were reported in 52 (100%) of the MAs.

4. *Bias in the selection of the studies*

The avoidance of bias in the selection of studies was reported in 13 (25%) of the MAs.

5. *Methodological quality*

The criteria used for assessing the validity of the included studies were reported in 18 (34.6%) of the MAs.

6. *Validity assessed appropriately*

Appropriate criteria for assessing internal or external validity were reported in 41 (78.8%) of the MAs.

7. *Methods used to combine the findings*

The methods used to combine the findings of included studies were reported in 32 (61.5%) of the MAs.

8. *Findings combined appropriately*

The findings of the relevant studies were combined appropriately relative to the primary question in 50 (96.2%) of the MAs.

9. Conclusions supported by data

The conclusions the authors reached were supported by the data they adduced in 43 (82.7%) of the MAs.

10. Overall Quality

The overall quality of the MAs published in The Cochrane Library (1996) received a mean score of 3.42 (95% CI 2.92, 3.93 – see Table 3).

Table 3

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview Quality Assessment Questionnaire Published in the Cochrane Database of Systematic Reviews Between 1993 and 1996

Question	Yes	Partially or can't tell	No
1. Were the search methods used to find evidence reported?	18 [34.6]	33 [63.5]	1 [1.9]
2. Was the search strategy for evidence reasonable comprehensive	14 [26.9]	35 [67.3]	3 [5.8]
3. Were the criteria used for deciding which studies to include in the overview reported	52 [100.0]	0	0
4. Was bias in the selection of studies avoided	13 [25.0]	25 [48.1]	14 [26.9]
5. Were the criteria used for assessing the validity of the included studies reported	18 [34.6]	30 [57.7]	4 [7.7]
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	41 [78.8]	8 [15.4]	3 [5.8]
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	32 [61.5]	9 [17.3]	11 [21.2]
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	50 [96.2]	1 [1.9]	1 [1.9]
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	43 [82.7]	9 [17.3]	0
10. How would you rate the scientific quality of this overview?	3.42 (95% CI 2.92, 3.93)		

* See Appendix C for a full description of each item

3.2.4 Comparing paper-based MAs published in the ‘Early Years’ with those published in the ‘Later Years’

The MAs from the ‘Early Years’ were compared with those from the ‘Later Years’ with respect to the percentage of ‘yes’ ratings they received on questions 1 to 9 of the Oxman and Guyatt Scale (1991). MAs of the ‘Later Years’ received a higher percentage of ‘yes’ answers on eight questions. MAs of the ‘Earlier Years’ received a marginally higher percentage of ‘yes’ answers on one question. On average, the MAs of the ‘Later Years’ received 6.8% more ‘yes’ answers than those of the ‘Early Years’ (see Table 4).

Table 4

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview
Quality Assessment Questionnaire: Comparisons of Paper-based Data Published Between 1977 and 1989 and
Paper-based Data Published Between 1990 and 1995

Question	Published < 1990 Yes	Published 1990- 1995 Yes	Difference in % Response
1. Were the search methods used to find evidence reported?	28 [59.6] [95% CI 45.0, 74.1]	35 [67.3] [95% CI 54.1, 80.5]	7.73 [95% CI -12.0, 27.0]
2. Was the search strategy for evidence reasonable comprehensive	16 [34.0] [95% CI 20.0, 48.1]	22 [42.3] [95% CI 28.4, 56.2]	8.27 [95% CI -11.0, 28.0]
3. Were the criteria used for deciding which studies to include in the overview reported	27 [57.4] [95% CI 42.8, 72.1]	38 [73.1] [95% CI 60.6, 85.6]	15.7 [95% CI -3.27, 35.0]
4. Was bias in the selection of studies avoided	9 [19.1] [95% CI 07.5, 30.8]	11 [21.2] [95% CI 09.7, 32.6]	2.1 [95% CI -14.0, 18.0]
5. Were the criteria used for assessing the validity of the included studies reported	15 [31.9] [95% CI 18.1, 45.8]	25 [48.1] [95% CI 34.0, 62.1]	16.2 [95% CI -3.37, 36.0]
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	40 [85.1] [95% CI 74.5, 95.7]	47 [90.4] [95% CI 82.1, 98.7]	5.3 [95% CI -7.85, 18]
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	37 [78.7] [95% CI 66.6, 90.9]	45 [86.5] [95% CI 76.9, 96.1]	7.8 [95% CI -7.32, 23]
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	44 [93.6] [95% CI 86.4, 101]	47 [90.4] [95% CI 82.1, 98.7]	-3.2 [95% CI -14.0, 7.75]
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	40 [85.1] [95% CI 74.5, 95.7]	45 [86.5] [95% CI 76.9, 96.1]	1.4 [95% CI -13, 15]
10. How would you rate the scientific quality of this overview?	3.04 [95% CI 2.53, 3.56]	3.35 [95% CI 2.83, 3.87]	0.31 [95% CI -42, 1.03]

*See Appendix C for a full description of each item

For the following two questions, the difference in percentage of 'yes' answers obtained for these two groups of MA's was greater than 10%.

1. Criteria for inclusion/exclusion (Question 3)

Criteria used for deciding which studies to include in the MA were reported in 57.4% of MAs from the 'Early Years' compared to 73.1% of those from the 'Later Years'. This amounts to a difference of 15.7% (95% CI -3.27, 35.0).

2. Methodological quality (Question 5)

Criteria used for assessing the validity of included studies were reported in 31.9% of MAs from the 'Early Years' compared to 48.1% of those from the 'Later Years'. This amounts to a difference of 16.2% (95% CI -3.37, 36.0).

On question ten, which assessed the overall quality of reporting, MAs of the 'Later Years' received a mean score of 3.35 (95% CI 2.83, 3.87), while papers of the 'Early Years' received a mean score of 3.04 (95% CI 2.53, 3.56), with a mean difference of 0.31% (95% CI -.42, 1.03).

3.2.5 Comparing paper-based MAs and electronically published MAs

Paper-based MAs were compared with *electronically published* MAs with respect to the percentage of 'yes' ratings they received on questions 1 to 9 of the Oxman and Guyatt Scale (1991). Paper-based MAs received a higher percentage of 'yes' answers on six questions. The average percentage of difference on these questions was 17%. Electronically published MAs received a higher percentage of 'yes' answers on three questions. The average percentage of difference on these questions was 12.2% (see Table 5).

Table 5

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview Quality Assessment Questionnaire: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996

Question	Published Later Yes	Cochrane Yes	% Difference
1. Were the search methods used to find evidence reported?	35 [67.3] [95% CI 54.1, 80.5]	18 [34.6] [95% CI 21.2, 48.0]	-32.7 [95% CI -51.0, -14.0]
2. Was the search strategy for evidence reasonable comprehensive	22 [42.3] [95% CI 28.4, 56.2]	14 [26.9] [95% CI 14.5, 39.4]	-15.4 [95% CI -34.0, 3.6]
3. Were the criteria used for deciding which studies to include in the overview reported	38 [73.1] [95% CI 60.6, 85.6]	52 [100.0]	26.9 [95% CI 15.0, 39.0]
4. Was bias in the selection of studies avoided	11 [21.2] [95% CI 09.7, 32.6]	13 [25.0] [95% CI 12.8, 37.2]	3.8 [95% CI -13.0, 20.0]
5. Were the criteria used for assessing the validity of the included studies reported	25 [48.1] [95% CI 34.0, 62.1]	18 [34.6] [95% CI 21.2, 48.0]	-13.5 [95% CI -33.0, 5.7]
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	47 [90.4] [95% CI 82.1, 98.7]	41 [78.8] [95% CI 67.4, 90.3]	-11.6 [95% CI -26.0, 2.5]
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	45 [86.5] [95% CI 76.9, 96.1]	32 [61.5] [95% CI 47.9, 75.2]	-25 [95% CI -42.0, 8.5]
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	47 [90.4] 95% CI [82.1, 98.7]	50 [96.2] [95% CI 90.8, 102]	5.8 [95% CI -4.01, 16]
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	45 [86.5] [95% CI 76.9, 96.1]	43 [82.7] [95% CI 72.1, 93.3]	-3.8 [95% CI -18, 10]
10. How would you rate the scientific quality of this overview?	3.35 [95% CI 2.83, 3.87]	3.42 [95% CI 2.92, 3.93]	.07 [95% CI -.64,.79]

*See Appendix C for a full description of each item

On the following questions, the difference in percentage of 'yes' answers obtained for these two groups of MAs was greater than a 10%.

1. Search methods (Question 1)

Search methods used to find evidence were reported in 67.3% of the paper-based MAs, compared to 34.6% of those electronically published. This amounts to a difference of -32.7% (95% CI -51.0, -14.0).

2. Comprehensive search (Question 2)

A reasonably comprehensive search strategy was reported in 42.3% of the paper-based MAs, compared to 26.9% of those electronically published. This amounts to a difference of -15.4% (95% CI -34.0, -3.6).

3. Criteria for inclusion/exclusion (Question 3)

Criteria used for deciding which studies to include in the MA were reported in 73.1% of the paper-based MAs, compared to 100% of those electronically published. This amounts to a difference of 26.9% (95% CI 15,39).

4. Methodological quality (Question 5)

Criteria used for assessing the validity of the included studies were reported in 48.1% of the paper-based MAs, compared to 34.6% of those electronically published. This amounts to a difference of -13.5% (95% CI -33.0, 5.7).

5. Validity assessed appropriately (Question 6)

Appropriate criteria for assessing the validity of all included studies were reported in 90.4% of the paper-based MAs, compared to 78.8% of those electronically published. This amounts to a difference of -11.6% (95% CI -26.0, -2.5).

6. Methods used to combine the findings (Question 7)

The methods used to combine the findings of the relevant studies were reported in 86.5% of the paper-based MAs, compared to 61.5% of those electronically published. This amounts to a difference of -25 (95% CI -42, 8.5).

On question ten, which assessed the overall quality of reporting, the mean score was slightly higher for the electronically published MAs (3.42) than for the paper-published MAs (3.35), with a mean difference of 0.07 (95% CI -0.64, 0.79) (Figures 3 and 4).

Figure 3

Average Scores (Overall Q10, 1 to 7) for 'Early' and the 'Later' Years and Electronically Published (Cochrane) Meta-analyses

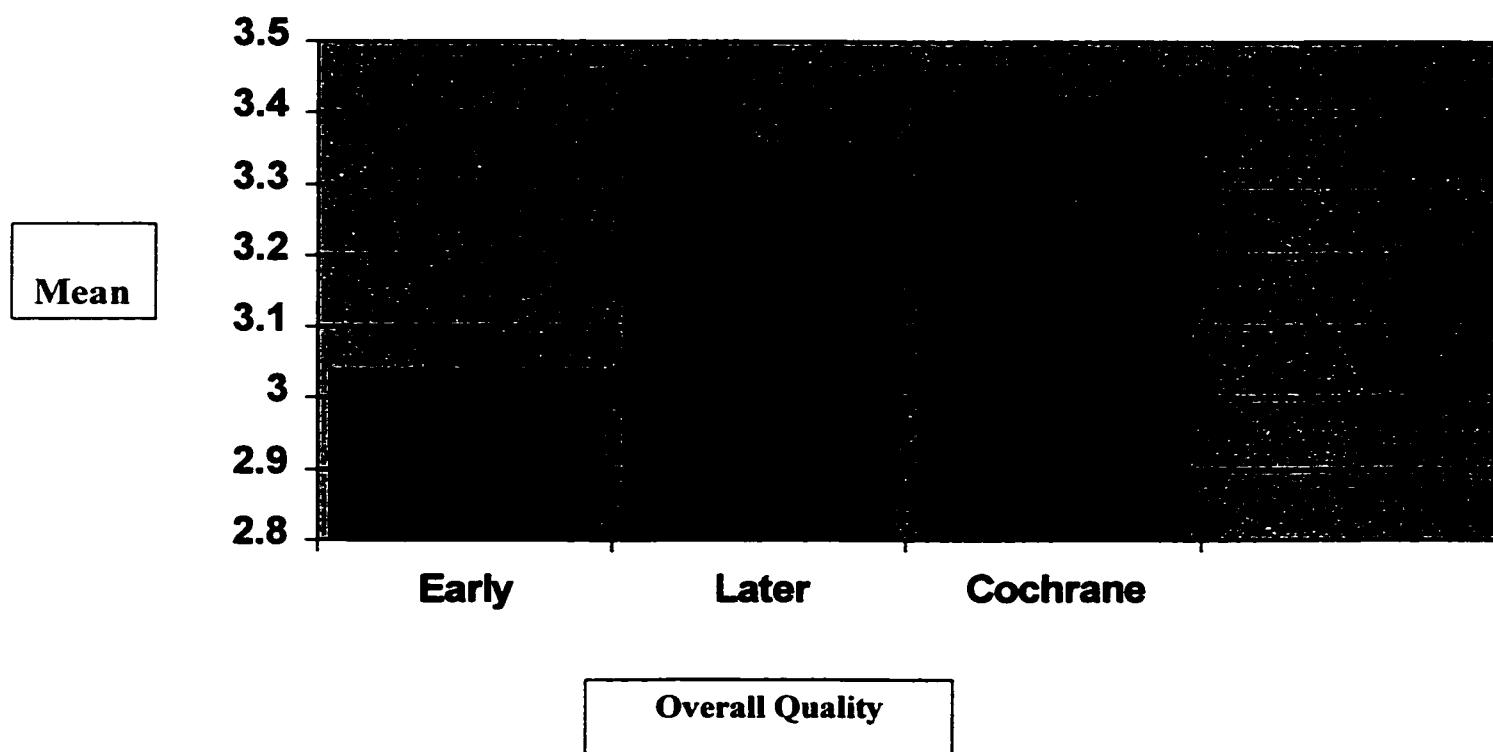
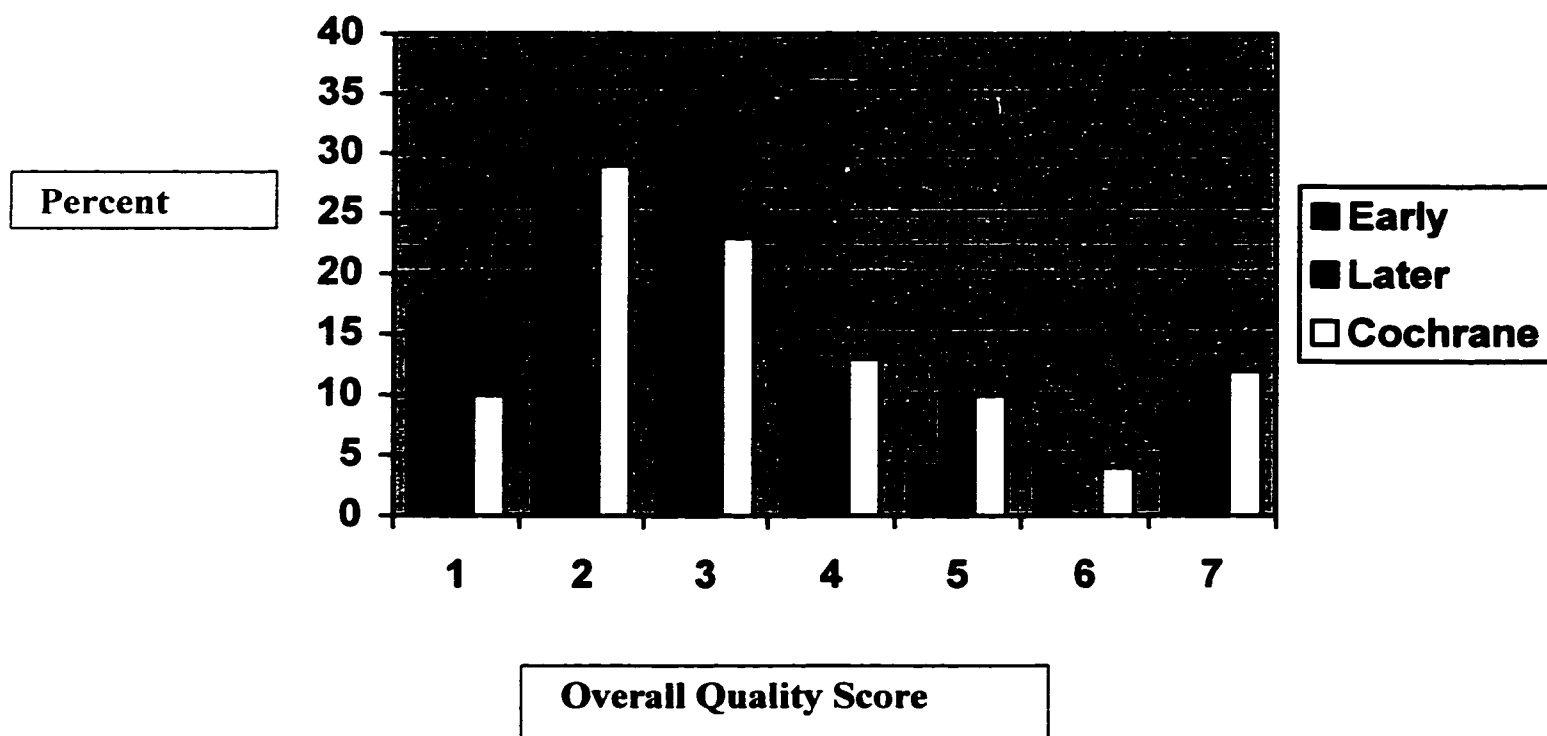


Figure 4

Paper-based Published Versus Electronically Published (Cochrane)



3.3 *Sacks et al. Checklist (1987) Results*

3.3.1 *Paper-based published MAs from the 'Early Years'*

1. *Prospective design*

Six (12.8%) MAs from the 'Early Years' reported that an 'a priori' protocol was established. Thirty (63.8%) MAs reported that the authors relied on more than just the literature search from MEDLINE. Thirty-seven (78.3%) listed the trials that were included in the analysis. Twelve (25.5%) listed the trials rejected from the MA. Twenty-nine (61.7%) listed the treatment assignments. Eleven (23.4%) provided the range of patients. Twenty-six (55.3%) provided the range of treatment. Eighteen (38.3%) listed the range of diagnosis.

2. *Combinability*

Twelve MAs (25.5%) published a statement of the criteria used for deciding whether or not the trials studied were similar enough to be pooled. Twenty-nine (61.7%) performed a test to assess homogeneity.

3. *Control of bias*

Eight MAs (17%) reported that the decision to include a paper was made by reviewing only the methods section of the trial and not the results, or that blinding was performed and the two areas were reviewed independently. Six of the reports (12.8%) stated that more than one observer extracted the data from the trials, and stated whether they were blinded using a photocopying process. Seven of the MAs reported that they assessed the inter-observer agreement among the reviewers. One MA reported the source of financial support and clearly acknowledged a potential conflict of interests.

4. *Statistical analysis*

Thirty-nine (83.0%) of the MAs reported using a recognized method of pooling. Eight (17.0%) mentioned awareness of potential type I or type II errors. Thirty-seven (78.7%) included confidence intervals for their significant results. Twenty-three (48.9%) carried out subgroup analysis to increase statistical power.

5. Sensitivity analysis

Twelve (25.5%) reported the methodological rigor and scientific quality of papers and considered this in formulating recommendations. Six (12.8%) included a sensitivity analysis showing how results varied with the use of different assumptions, tests and criteria.

Publication bias

Eight (17.0%) of the MAs reported using a method for calculating the number of unpublished negative studies that would be required to refute the published evidence.

6. Application of results

Twenty-eight (59.6%) of the MAs attempted to place the results of the MA into a perspective based on the overall criteria used to carry out the MA. One (2.1%) of the MAs reported the economic impact of results.

7. Language

Two (4.3%) of the MAs included studies in all languages (Table 6).

Table 6

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Questionnaire Published in Paper-based Peer Review Journals Between 1977 and 1989

Questions	Adequate	Partial	None or Unknown
1. Prospective Design			
a. Protocol	6 [12.8]	6 [12.8]	35 [74.5]
b. Literature search	30 [63.8]	9 [19.1]	8 [17.0]
c. Lists of trials analyzed	37 [78.7]	7 [14.9]	3 [6.4]
d. Log of rejected trials	12 [25.5]	3 [6.4]	32 [68.1]
e. Treatment assignment	29 [61.7]	10 [21.3]	8 [17.0]
f. Ranges of patients	11 [23.4]	17 [36.2]	19 [40.4]
g. Ranges of treatment	26 [55.3]	9 [19.1]	12 [25.5]
h. Ranges of diagnosis	18 [38.3]	11 [23.4]	18 [38.3]
2. Combinability			
a. Criteria	12 [25.5]	16 [34.0]	19 [40.4]
b. Measurement	29 [61.7]	4 [8.5]	14 [29.8]
3. Control of bias			
a. Selection bias	8 [17.0]	4 [8.5]	35 [74.5]
b. Data-extraction bias	6 [12.8]	2 [4.3]	39 [83.0]
c. Inter-observer agreement	7 [14.9]	1 [2.1]	39 [83.0]
d. Source of support	1 [2.1]	1 [2.1]	45 [95.7]
4. Statistical analysis			
a. Statistical methods	39 [83.0]	7 [14.9]	1 [2.1]
b. Statistical errors	8 [17.0]	3 [6.4]	36 [76.6]
c. Confidence intervals	37 [78.7]	1 [2.1]	9 [19.1]
d. Subgroup analysis	23 [48.9]	10 [21.3]	14 [29.8]
5. Sensitivity Analysis			
a. Quality assessment	12 [25.5]	5 [10.6]	30 [63.8]
b. Varying methods	6 [12.8]	7 [14.9]	34 [72.3]
c. Publication bias	8 [17.0]	11 [23.4]	28 [59.6]
6. Application of results			
a. Caveats	28 [59.6]	18 [38.3]	1 [2.1]
b. Economic Impact	1 [2.1]	0	46 [97.9]
7. Language	2 [4.3]	1 [2.1]	44 [93.6]

*See Appendix G and H for a full description of each item

3.3.2 Paper-based published MAs from the 'Later Years'

1. Prospective design

Three (5.8%) of MAs of the 'Early Years' reported that an 'a priori' protocol was established. Thirty-three (63.5%) reported that they relied on more than just the literature search from MEDLINE. Forty one (78.8%) listed the trials that were included in the analysis and 17 (32.7%) listed the trials rejected from the MA. Thirty-three (63.5%) listed the treatment assignments; nineteen (36.5%) provided the range of patients; thirty-two (61.5%) provided the range of treatment; and twenty-two (42.3%) listed the range of diagnosis.

2. Combinability

Fifteen MAs (28.8%) published a statement of the criteria used to decide whether the trials were similar enough to be pooled. Thirty-four (65.4%) performed a test to assess homogeneity.

3. Control of bias

Seven MAs (13.5%) reported that the decision to include a paper was made by reviewing only the methods section of the trial and not the results, or that blinding was performed and the two areas reviewed independently. Fourteen (26.9%) stated that more than one observer extracted the data from the trials and included whether they were blinded using a photocopying process. Seven (13.5%) reported that they assessed the inter-observer agreement among the reviewers. Three (5.8%) reported their source of financial support and clearly acknowledged potential conflict of interests.

4. Statistical analysis

Forty-four (84.6%) of the MAs reported using a recognized pooling method. Six (11.5%) mentioned the awareness of potential type I or type II errors. Forty-one (78.8%) included confidence intervals for their significant results. Thirty (57.7%) carried out subgroup analysis to increase statistical power.

5. Sensitivity analysis

Twenty-two (42.3%) reported the methodological rigor and scientific quality of papers and

considered this in formulating recommendations. Thirteen (25.0%) included a sensitivity analysis showing how the results varied with the use of different assumptions, tests and criteria.

Publication bias

Nine (17.3%) of the MAs reported using a method to calculate the number of unpublished negative studies that would be required to refute the published evidence.

6. Application of results

Thirty-six (69.2%) of the MAs attempted to place their results into a perspective based on the overall criteria used to carry out the MA. Three (5.8%) reported the economic impact of their results.

7. Language

Two (3.8%) of the MAs included studies in all languages (Table 7).

Table 7

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Questionnaire Published in Peer Review Journals Between 1990 and 1995

Questions	Adequate	Partial	None or Unknown
1. Prospective Design			
a. Protocol	3 [5.8]	7 [13.5]	42 [80.8]
b. Literature search	33 [63.5]	9 [17.3]	10 [19.2]
c. Lists of trials analyzed	41 [78.8]	6 [11.5]	5 [9.6]
d. Log of rejected trials	17 [32.7]	7 [13.5]	28 [53.8]
e. Treatment assignment	33 [63.5]	4 [7.7]	15 [28.8]
f. Ranges of patients	19 [36.5]	4 [7.7]	29 [55.8]
g. Ranges of treatment	32 [61.5]	7 [13.5]	13 [25.0]
h. Ranges of diagnosis	22 [42.3]	11 [21.2]	19 [36.5]
2. Combinability			
a. Criteria	15 [28.8]	14 [26.9]	23 [44.2]
b. Measurement	34 [65.4]	7 [13.5]	11 [21.2]
3. Control of bias			
a. Selection bias	7 [13.5]	6 [11.5]	39 [75.0]
b. Data-extraction bias	14 [26.9]	7 [13.5]	31 [59.6]
c. Inter-observer agreement	7 [13.5]	5 [9.6]	40 [76.9]
d. Source of support	3 [5.8]	1 [1.9]	48 [92.3]
4. Statistical analysis			
a. Statistical methods	44 [84.6]	7 [13.5]	1 [1.9]
b. Statistical errors	6 [11.5]	4 [7.7]	42 [80.8]
c. Confidence intervals	41 [78.8]	1 [1.9]	10 [19.2]
d. Subgroup analysis	30 [57.7]	6 [11.5]	16 [30.8]
5. Sensitivity Analysis			
a. Quality assessment	22 [42.3]	7 [13.5]	23 [44.2]
b. Varying methods	13 [25.0]	6 [11.5]	33 [63.5]
c. Publication bias	9 [17.3]	11 [21.2]	32 [61.5]
6. Application of results			
a. Caveats	36 [69.2]	13 [25.0]	3 [5.8]
b. Economic Impact	3 [5.8]	2 [3.8]	47 [90.4]
7. Language	2 [3.8]	2 [3.8]	48 [92.3]

*See Appendix G and H for a full description of each item

3.3.3 Electronically published MAs from The Cochrane Library

1. Prospective design

Two (3.8%) Cochrane Library MAs reported that an 'a priori' protocol was established. Sixteen (30.8%) reported that they relied on more than just the literature search from MEDLINE. Fifty (96.2%) listed the trials that were included in the analysis and thirty-one (59.6%) listed the trials rejected from the MA. Forty-one (94.2%) listed the treatment assignments; seventeen (32.7%) provided the range of patients; fifty (96.2%) provided the range of treatment and thirty-eight (73.1%) listed the range of diagnoses.

2. Combinability

Ten MAs (19.2%) published a statement of the criteria used to decide whether the trials were similar enough to be pooled. Fifteen (28.8%) performed a test to assess homogeneity.

3. Control of bias

Ten MAs (19.2%) reported that the decision to include a paper was made by reviewing only the methods section of the trial and not the results, or that blinding was performed and the two areas reviewed independently. Sixteen (30.8%) stated that more than one observer extracted the data from the trials, and included whether they were blinded using a photocopying process. One (1.9%) reported that it assessed the inter-observer agreement among the reviewers. One reported its source of financial support and clearly acknowledged potential conflict of interests.

4. Statistical analysis

Fifty-one (98.1%) MAs reported using a recognized method of pooling. Two (3.8%) mentioned awareness of potential type I or type II errors. Forty-eight (92.3%) included confidence intervals for their significant results. Thirty-nine (75.0%) carried out subgroup analysis to increase statistical power.

5. Sensitivity analysis

Twenty-one (40.4%) MAs reported the methodological rigor and scientific quality of papers and considered this in formulating recommendations. Five (9.6%) included a sensitivity

analysis showing how results varied with the use of different assumptions, tests and criteria.

Publication bias

Four (7.7%) MAs reported that they used a method to calculate the number of unpublished negative studies that would be required to refute the published evidence.

6. Application of results

Twenty-seven (51.9%) MAs attempted to place their results into a perspective based on the overall criteria used to carry out the MA. One (1.9%) reported the economic impact of its results.

7. Language

Two (3.8%) MAs included studies in all languages (Table 8).

Table 8

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist Published in the Cochrane Database of Systematic of Reviews Between 1993 and 1996

Questions	Adequate	Partial	None or Unknown
1. Prospective Design			
a. Protocol	2 [3.8]	3 [5.8]	47 [90.4]
b. Literature search	16 [30.8]	34 [65.4]	2 [3.8]
c. Lists of trials analyzed	50 [96.2]	0	2 [3.8]
d. Log of rejected trials	31 [59.6]	1 [1.9]	20 [38.5]
e. Treatment assignment	49 [94.2]	1 [1.9]	2 [3.8]
f. Ranges of patients	17 [32.7]	8 [15.4]	27 [51.9]
g. Ranges of treatment	50 [96.2]	0	2 [3.8]
h. Ranges of diagnosis	38 [73.1]	11 [21.2]	3 [5.8]
2. Combinability			
a. Criteria	10 [19.2]	8 [15.4]	34 [65.4]
b. Measurement	15 [28.8]	3 [5.8]	34 [65.4]
3. Control of bias			
a. Selection bias	10 [19.2]	2 [3.8]	40 [76.9]
b. Data-extraction bias	16 [30.8]	1 [1.9]	35 [67.3]
c. Inter-observer agreement	1 [1.9]	8 [15.4]	43 [82.7]
d. Source of support	45 [86.5]	0	7 [13.5]
4. Statistical analysis			
a. Statistical methods	51 [98.1]	0	1 [1.9]
b. Statistical errors	2 [3.8]	1 [1.9]	49 [94.2]
c. Confidence intervals	48 [92.3]	0	4 [7.7]
d. Subgroup analysis	39 [75.0]	7 [13.5]	6 [11.5]
5. Sensitivity Analysis			
a. Quality assessment	21 [40.4]	31 [59.6]	0
b. Varying methods	5 [9.6]	3 [5.8]	44 [84.6]
c. Publication bias	4 [7.7]	2 [3.8]	46 [88.5]
6. Application of results			
a. Caveats	27 [51.9]	23 [44.2]	2 [3.8]
b. Economic Impact	1 [1.9]	6 [11.5]	45 [86.5]
7. Language	2 [3.8]	1 [1.9]	49 [94.2]

*See Appendix G and H for a full description of each item

3.3.4 Comparing paper-based published MAs from the ‘Early Years’ with those of the ‘Later Years’

The following questions generated a difference greater than 10%.

1. Prospective design

Eleven from the ‘Early Years’ (23.4% (95% CI 10.9%, 36.0%)) compared with 19 from the ‘Later Years’ (36.5% (95% CI 23.0%, 50.1%)) gave the range of patients.

2. Control of bias

Six of the trials (12.8% (95% CI 2.9%, 22.7%)) from the ‘Early Years’ compared with 14 (26.9% (95% CI 14.5%, 39.4%)) from the ‘Later Years’ reported that more than one observer extracted the data from the trials, and included whether they were blinded using a photocopying process.

3. Sensitivity analysis

Twelve (25.5% (95% CI 12.6%, 38.5%)) compared with twenty-two (42.3% (95% CI 28.4%, 56.2%)) reported the methodological rigor and scientific quality of papers and considered this in formulating recommendations. Six (12.8% (95% CI 2.8%, 22.7%)) compared with 13 (25.0% (95% CI 12.8%, 37.2%)) included a sensitivity analysis showing how results varied with the use of different assumptions, tests and criteria.

4. Publication bias

No difference was found in publication bias.

5. Language

No difference was found in the language of included studies.

On seventeen out of the twenty-four items (70%), the paper-based MAs from the ‘Later Years’ received higher scores than those from the ‘Early Years’ (Table 9).

Table 9

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist: Comparisons of Paper-based Data Published Between 1977 and 1989 and Paper-based Data Published Between 1990 and 1995

Questions	Adequate (1977-1989)	Adequate (1990-1995)	% DIFFERENCE
1. Prospective Design			
a. Protocol	6 [12.8] [95% CI 02.9, 22.7]	3 [5.8] [95% CI -00.8, 12.3]	-7.0 [95% CI -19, 4.8]
b. Literature search	30 [63.8] [95% CI 49.6, 78.1]	33 [63.5] [95% CI 49.9, 77.0]	-0.3 [95% CI -20, 19]
c. Lists of trials analyzed	37 [78.7] [95% CI 66.6, 90.9]	41 [78.8] [95% CI 67.4, 90.3]	0.1 [95% CI -16, 17]
d. Log of rejected trials	12 [25.5] [95% CI 12.6, 38.5]	17 [32.7] [95% CI 19.5, 45.9]	7.2 [95% CI -11, 25]
e. Treatment assignment	29 [61.7] [95% CI 47.3, 76.1]	33 [63.5] [95% CI 49.9, 77.0]	1.8 [95% CI -18, 21]
f. Ranges of patients	11 [23.4] [95% CI 10.8, 36.0]	19 [36.5] [95% CI 23.0, 50.1]	13.1 [95% CI -5.2, 31]
g. Ranges of treatment	26 [55.3] [95% CI 40.6, 70.1]	32 [61.5] [95% CI 47.9, 75.2]	6.2 [95% CI -14, 26]
h. Ranges of diagnosis	18 [38.3] [95% CI 23.9, 52.7]	22 [42.3] [95% CI 28.4, 56.2]	4.0 [95% CI -16, 24]
2. Combinability			
a. Criteria	12 [25.5] [95% CI 12.6, 38.5]	15 [28.8] [95% CI 16.1, 41.6]	3.3 [95% CI -15, 21]
b. Measurement	29 [61.7] [95% CI 47.3, 76.1]	34 [65.4] [95% CI 52.0, 78.8]	3.7 [95% CI -16, 23]
3. Control of bias			
a. Selection bias	8 [17.0] [95% CI 05.9, 28.2]	7 [13.5] [95% CI 03.9, 23.1]	-3.5 [95% CI -18, 11]
b. Data-extraction bias	6 [12.8] [95% CI 02.9, 22.7]	14 [26.9] [95% CI 14.5, 39.4]	14.1 [95% CI -1.6, 30]
c. Inter-observer agreement	7 [14.9] [95% CI 04.3, 25.5]	7 [13.5] [95% CI 03.9, 23.1]	-1.4 [95% CI -16, 13]
d. Source of support	1 [2.1] [95% CI -02.2, 06.4]	3 [5.8] [95% CI -00.8, 12.3]	3.7 [95% CI -4.1, 11]
4. Statistical analysis			
a. Statistical methods	39 [83.0] [95% CI 71.8, 94.1]	44 [84.6] [95% CI 74.5, 94.8]	1.6 [95% CI -13, 16]
b. Statistical errors	8 [17.0] [95% CI 05.9, 28.2]	6 [11.5] [95% CI 02.6, 20.5]	-5.5 [95% CI -20, 8.7]
c. Confidence intervals	37 [78.7] [95% CI 66.6, 90.9]	41 [78.8] [95% CI 67.4, 90.3]	0.1 [95% CI -16, 17]
d. Subgroup analysis	24 [51.1] [95% CI 36.2, 65.9]	30 [57.7] [95% CI 43.8, 71.6]	6.6 [95% CI -13, 27]
5. Sensitivity Analysis			
a. Quality assessment	12 [25.5] [95% CI 12.6, 38.5]	22 [42.3] [95% CI 28.4, 56.2]	16.8 [95% CI -2.0, 36]
b. Varying methods	6 [12.8] [95% CI 02.9, 22.7]	13 [25.0] [95% CI 12.8, 37.2]	12.2 [95% CI -3.3, 28]
c. Publication bias	8 [17.0] [95% CI 05.9, 28.2]	9 [17.3] [95% CI 06.7, 27.9]	0.3 [95% CI -15, 16]

6. Application of results			
a. Caveats	28 [59.6] [95% CI 45.0, 74.1]	36 [69.2] [95% CI 56.3, 82.2]	9.6 [95% CI -9.6, 29]
b. Economic Impact	1 [2.1] [95% CI -02.2, 06.4]	3 [5.8] [95% CI 00.8, 12.3]	3.7 [95% CI -4.1, 11]
7. Language	2 [4.3] [95% CI -01.7, 10.3]	2 [3.8] [95% CI -01.6, 09.3]	-0.5 [95% CI -8.4, 7.6]

*See Appendix G and H for a full description of each item

3.3.5 Comparing 'Later Years' paper-based and (Cochrane) electronically published MAs

The following questions generated differences greater than 10%.

1. Prospective design

Thirty-three of the paper-based MAs (63.5% (95% CI 49.9%, 77.0%)) compared to sixteen (30.8% (95% CI 17.8%, 43.7%)) from the Cochrane database reported that they relied on more than just the literature search from MEDLINE. Forty-one (78.8% (95% CI 67.4%, 90.3%)) from paper-based and 50 (96.2% (95% CI 90.75%, 101.6%)) from the Cochrane database listed the trials that were included in the analysis. Seventeen (32.7% (95% CI 19.5%, 45.9%)) paper-based MAs compared to thirty-one (59.6% (95% CI 45.8%, 73.4%)) of the Cochrane database listed the trials rejected from the MA. Thirty-three of the paper-based MAs (63.5% (95% CI 49.9%, 77.0%)) compared to forty-nine (94.2% (95% CI 87.7%, 100.8%)) listed the treatment assignments. Twenty-two (61.5% (95% CI 47.9%, 75.2%)) compared with fifty (96.2% (95% CI 90.75%, 101.56%)) provided the range of treatment; and twenty-two (42.3% (95% CI 28.4%, 56.2%)) compared with thirty-eight (73.1% (95% CI 60.6%, 85.5%)) listed the range of diagnoses.

2. Combinability

Thirty-four (65.4% (95% CI 52.0%, 78.8%)) of the paper-based MAs and fifteen (28.8% (95% CI 16.1%, 41.6%)) from the Cochrane database performed a test to assess homogeneity.

3. Control of bias

Seven (13.5% (95% CI 03.9%, 23.1%)) of the paper-based MAs and 1 (1.9% (95% CI -1.9, 5.8%)) from the Cochrane database reported assessing the inter-observer agreement among reviewers. Three (5.8% (95% CI -0.8%, 12.3%)) of the paper-based MAs compared with forty-five (86.5% (95% CI 76.9%, 96.1%)) from the Cochrane database reported their source of financial support and clearly acknowledged a potential conflict of interests.

4. Statistical analysis

Forty-four (84.6% (95% CI 74.5%, 94.8%)) of the paper-based MAs and fifty-one (98.1% (95% CI-94.2%, 101.9%)) Cochrane Database MAs reported using a recognized pooling method. Forty-one (78.8% (95% CI 67.4%, 90.3%)) compared to forty-eight (92.3% (95% CI-84.8%, 99.9%)) included confidence intervals for their significant results. Thirty (57.7% (95% CI-43.8%, 71.6%)) of the paper-based MAs and thirty-nine (75.0% (95% CI 62.8%, 87.2%)) from the Cochrane Database conducted subgroup analysis in order to increase statistical power.

5. Sensitivity analysis

Thirteen (25.0% (95% CI 12.8%, 37.2%)), compared to five (9.6% (95% CI -0.8%, 12.3%)), included a sensitivity analysis showing how their results varied with the use of different assumptions, tests and criteria.

6. Application of results

Thirty-six (69.2% (95% CI 56.3%, 82.2%)), compared with twenty-seven (51.9% (95% CI 37.9%, 66.0%)), attempted to put their results into perspective based on the overall criteria used to carry out the MA.

7. Language

There was no difference in the inclusion of studies in languages other than English.

On eleven (45%) of the twenty-four items (including language), the Cochrane Database MAs received higher scores than the paper-based MAs of the 'Later Years' (Table 10).

Table 10

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist: Comparisons of Paper-based Data Between 1990 and 1995 and Cochrane Data Between 1993 and 1996

Questions	Adequate/Later	Adequate/Cochrane	% Difference
1. Prospective Design			
a. Protocol	3 [5.8] [95% CI 00.8, 12.3]	2 [3.8] [95% CI 01.6, 09.3]	-2.0 [95% CI -10. 6.5]
b. Literature search	33 [63.5] [95% CI 49.9, 77.0]	16 [30.8] [95% CI 17.8, 43.7]	-32.7 [95% CI -51. 14]
c. Lists of trials analyzed	41 [78.8] [95% CI 67.4, 90.3]	50 [96.2] [95% CI 90.8, 102]	17.4 [95% CI 4.7, 30]
d. Log of rejected trials	17 [32.7] [95% CI 19.5, 45.9]	31 [59.6] [95% CI 45.8, 73.4]	26.9 [95% CI 8.1, 46]
e. Treatment assignment	33 [63.5] [95% CI 49.9, 77.0]	49 [94.2] [95% CI 87.7, 101]	30.7 [95% CI 16. 46]
f. Ranges of patients	19 [36.5] [95% CI 23.0, 50.1]	17 [32.7] [95% CI 19.5, 45.9]	-3.8 [95% CI -23. 15]
g. Ranges of treatment	22 [61.5] [95% CI 47.9, 75.2]	50 [96.2] [95% CI 90.8, 102]	34.7 [95% CI 20. 49]
h. Ranges of diagnosis	22 [42.3] [95% CI 28.4, 56.2]	38 [73.1] [95% CI 60.6, 85.6]	30.8 [95% CI 12, 49]
2. Combinability			
a. Criteria	15 [28.8] [95% CI 16.1, 41.6]	10 [19.2] [95% CI 08.2, 30.3]	-9.6 [95% CI 12, 49]
b. Measurement	34 [65.4] [95% CI 52.0, 78.8]	15 [28.8] [95% CI 16.1, 41.6]	-35.6 [95% CI -26. 7.1]
3. Control of bias			
a. Selection bias	7 [13.5] [95% CI 03.9, 23.1]	10 [19.2] [95% CI 08.2, 30.3]	5.7 [95% CI -8.7, 20]
b. Data-extraction bias	14 [26.9] [95% CI 14.5, 39.4]	16 [30.8] [95% CI 17.8, 43.8]	3.9 [95% CI -14, 22]
c. Inter-observer agreement	7 [13.5] [95% CI 03.9, 23.1]	1 [1.9] [95% CI -01.9, 05.8]	-11.6 [95% CI -22, -1.3]
d. Source of support	3 [5.8] [95% CI -00.8, 12.3]	45 [86.5] [95% CI 76.9, 96.1]	80.7 [95% CI 69, 92]
4. Statistical analysis			
a. Statistical methods	44 [84.6] [95% CI 74.5, 94.8]	51 [98.1] [95% CI 94.2, 102]	13.5 [95% CI 2.7, 24]
b. Statistical errors	6 [11.5] [95% CI 02.6, 20.5]	2 [3.8] [95% CI -01.6, 09.3]	-7.7 [95% CI -18, 2.7]
c. Confidence intervals	41 [78.8] [95% CI 67.4, 90.3]	48 [92.3] [95% CI 84.8, 99.8]	13.5 [95% CI -8.3, 27]
d. Subgroup analysis	30 [57.7] [95% CI 43.8, 71.6]	39 [75.0] [95% CI 62.8, 87.2]	17.3 [95% CI -9.4, 36]
5. Sensitivity analysis			
a. Quality assessment	22 [42.3] [95% CI 28.4, 56.2]	21 [40.4] [95% CI 26.6, 54.2]	-1.9 [95% CI -21, 17]
b. Varying methods	13 [25.0] [95% CI 12.8, 37.2]	5 [9.6] [95% CI 01.3, 17.9]	-15.4 [95% CI -30, -8.4]
c. Publication bias	9 [17.3] [95% CI 06.7, 27.9]	4 [7.7] [95% CI 00.2, 15.2]	-9.6 [95% CI -22, 3.2]

6. Application of results			
a. Caveats	36 [69.2] [95% CI 56.3, 82.2]	27 [51.9] [95% CI 37.9, 66.0]	-17.3 [95% CI -36, 1.6]
b. Economic Impact	3 [5.8] [95% CI -00.8, 12.3]	1 [1.9] [95% CI -01.9, 05.8]	-3.9 [95% CI -11, 3.7]
7. Language	2 [3.8] [95% CI -01.6, 09.3]	2 [3.8] [95% CI -01.6, 09.3]	0 [95% CI -7.6, 7.6]

*See Appendix G and H for a full description of each item

3.4 Multivariate analysis (for all statistically significant and substantively significant greater than 10%)

3.4.1 Oxman and Guyatt - Year of Publication

All attributes were compared to the reference case, that is, to those MAs published between 1990 and 1995. Only those reaching statistical significance are reported below. The year of publication and number of component studies were included in the model as predictors, and are reported using levels of significance (i.e., 'p' values).

Relative to the paper-published MAs of the 'Later Years', MAs published in The Cochrane Library are less likely to report the search methods used to find evidence (odds ratio .339 (95% CI .145%, .798%)). The year of publication was a predictor for reporting the search methods used to find the evidence ($p=0.046$). In addition, the number of component studies was related to the reporting of the search strategy ($p=0.029$).

The paper-published MAs of the 'Later Years', compared to those published in The Cochrane Library were less likely to report the methods used to combine the findings of the relevant studies to reach a conclusion (odds ratio .324 (95% CI .118%, .891%)). The year of publication was not a predictor for reporting the methods used to combine the findings of the relevant studies to reach a conclusion ($p=0.092$). As well, the number of component studies was not related to the reporting of the methods used to combine the findings of the relevant studies to reach a conclusion ($p=0.081$).

3.4.2 Sacks - Year of Publication

Relative to the paper-based published of the 'Later Years', those published earlier were more likely to report the protocol odds ratio 2.4 (95% CI 0.6, 10.2). Similarly, those published in The Cochrane Library were slightly less likely to report the protocol, odds ratio 0.7 (0.1, 4.4). However, year of publication was not a predictor for reporting the protocol ($p = 0.27$). In addition, the number of component studies was not related to the reporting of the protocol ($p = 0.99$). The results of the Hosmer and Lemeshow Goodness-of-Fit Test were similar throughout and ranged from .56 to .68. (Appendix I)

Relative to the paper-published MAs of the 'Later Years', those published in The Cochrane Library were less likely to report the details of their search strategy (odds ratio .312 (95% CI .133%, .735%)). The year of publication was a predictor for reporting the complete details of the search strategy ($p=0.015$); but the number of component studies was not related to the reporting of the search strategy ($p=0.129$).

Relative to the paper-published MAs of the 'Later Years', those published in The Cochrane Library were more likely to report the complete list of trials analysed and the log of rejected trials (odds ratio 6.16 (95% CI 1.23%, 30.69%)). The year of publication was not a predictor for reporting the complete list of trials analysed and the log of rejected trials ($p=0.073$). As well, the number of component studies was not related to the reporting of the list of trials analysed and the log of rejected trials ($p=0.652$).

In comparison to the paper-published MAs of the 'Later Years', those published in The Cochrane Library were more likely to report the method of treatment assignment in the primary study (odds ratio 4.40 (95% CI 1.77%, 10.95%)). The year of publication was a predictor for reporting the method of treatment assignment in the primary study ($p=0.001$). As well, the number of component studies was also related to the reporting of the method of treatment assignment in the primary study ($p=0.045$).

In comparison to the paper-published MAs of the 'Later Years', the Cochrane reports were more likely to report the ranges of patient characteristics, diagnosis, and treatment (odds ratio 9.39 (95% CI 2.48%, 35.56%)). The year of publication was a predictor for reporting the ranges of patient characteristics, diagnosis, and treatment ($p=0.002$). The number of component studies was not related to the reporting of the ranges of patient characteristics, diagnosis, and treatment ($p=0.996$).

Relative to the paper-based MAs of the 'Later Years', Cochrane MAs were more likely to report their source of support (odds ratio 135.50 (95% CI 25.35%, 724.20%)). The year of publication was a highly significant predictor for reporting the source of support ($p=0.001$), but the number of component studies was not related to the reporting of the source of support

($p=0.514$).

Relative to the paper-published MAs of the 'Later Years', Cochrane MAs were more likely to report appropriate statistical methods (odds ratio 10.26 (95% CI 1.19%, 88.47%)). The year of publication was not a predictor for reporting appropriate statistical methods ($p=0.079$). Neither was the number of component studies related to the reporting of appropriate statistical methods ($p=0.617$).

In comparison to the paper-published MAs of the 'Later Years', Cochrane MAs were more likely to report a subgroup analysis (odds ratio 2.87 (95% CI 1.17%, 7.06%)). The year of publication was a highly significant predictor for reporting that a subgroup analysis was conducted ($p=0.012$), although, the number of component studies was not related to the reporting of a subgroup analysis ($p=0.082$).

3.5 Comparison to Sacks et al. (1987) data

It was virtually impossible to compare MAs of the same year to one another, because of the way Sacks et al. (1987) reported their results. In order to compare results that were representative of the sample years which Sacks et al. reported, we chose to compare the 'Early Years' (1977 to 1989) sample of MAs selected for this study with a similar sample from the Sacks et al data 1983 to 1987. The majority of MAs from both of these groups, are from the 1983 to 1987 time period.

On fourteen out of twenty-four items (including language) (58%), MAs of these two groups received a difference in average score of greater than ten percent. On all but one item, the 'Early Years' data set reported a lower mean and percentage than the data set of Sacks et al. Combinability, was the only item on which the 'Early Years' sample scored higher than the Sacks data set. Some of the largest differences (Table 11) were in the reporting of ranges of patients (38.6%); ranges of diagnosis (20.7%); criteria (25.5%); subgroup analyses (29.1%); varying methods (30.2%); publication bias (24.0%); and source of support (25.9%).

Table 11

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist: Comparison of Sacks Data Published Between 1983 and 1987 and Data Included in this Study as Published Between 1977 and 1989

Question	Published Sacks Yes [no. %]	Paper-published Shea Yes [no. %]	% Difference [% Difference]
1. Prospective Design			
Protocol	13 [22]	6 [12.8]	9.2
Literature	40 [69]	30 [63.8]	5.2*
List of trials analyzed	54 [93]	37 [78.7]	14.3*
Log of rejected trials	24 [41]	12 [25.5]	15.5*
Treatment assignment	46 [79]	29 [61.7]	17.3*
Ranges of patients	36 [62]	11 [23.4]	38.6*
Ranges of treatments	39 [67]	26 [55.3]	11.7*
Ranges of diagnosis	34 [59]	18 [38.3]	20.7*
2. Combinability			
Criteria	39 [67]	12 [25.5]	25.5*
Measurement	27 [47]	29 [61.7]	14.7*
3. Control of Bias			
Selection bias	7 (12)	8 [17.0]	-5.0*
Data-extraction Bias	7 (12)	6 [12.8]	-0.8*
Inter-observer agreement	11 [19]	7 [14.9]	4.1
Source of support	16 [28]	1 [2.1]	25.9*
4. Statistical analysis			
Statistical methods	45 [78]	39 [83.0]	-5.0*
Statistical errors	38 [66]	8 [17.0]	4.9*
Confidence intervals	49 [84]	37 [78.7]	5.3*
Subgroup analyses	45 [78]	23 [48.9]	29.1*
5. Sensitivity analysis			
Quality assessment	15 [26]	12 [25.5]	0.5
Varying methods	25 [43]	6 [12.8]	30.2*
Publication bias	24 [41]	8 [17.0]	24.0*
6. Application of results			
Caveats	41 [71]	28 [59.6]	11.4*
Economic impact	3 [5]	1 [2.1]	2.9
7. Language	2 [4.3]	2 [4.3]	

*See Appendix G and H for a full description of each item

3.6 *Summary of Results*

In summary, Cochrane MAs scored slightly higher than paper-based MAs of the 'Later Years'. They scored higher by more than 10% in their reporting of the following items: criteria used for including studies; lists of trials analysed; log of rejected trials; treatment assignment; ranges of patients; ranges of treatment; ranges of diagnosis; and source of support. Notably, they scored 80.7% higher on sources of support. Cochrane MAs scored higher on questions dealing with analytical factors and sensitivity of discrimination, including the reporting of statistical methods and errors; confidence intervals; and whether a subgroup analysis was conducted and reported.

Paper-based MAs were more likely to report assessing the quality of reports. They scored over 10% higher on the reporting of the following items: literature search; combinability (i.e., reporting of measurement, as well as the methods used for combining studies); inter-observer agreement; whether a sensitivity analysis was done, including how results vary with the use of different assumptions, tests and criteria; caveats concerning application of results; and application validity assessment.

Paper-based 'Early Years' versus 'Later Years' MAs

The measures provided by the quality assessment checklist and scale indicated an improvement over time in a number of areas. There was an improvement of 15.7% in the reporting of the criteria used to decide which studies to include in the MA (Oxman and Guyatt scale). There was also a 16.2% improvement in the reporting of the criteria used to assess the validity of the included studies. However, this difference was not statistically significant ($P=0.21$).

The data yielded by the checklist of Sacks et al. (1987) also indicated a general trend toward improvement in reporting over time. However, not all of these differences were statistically or substantively significant. Areas where the improvement observed was greater than 10% included items on: control of bias and sensitivity analysis; whether the data extraction was reported as involving more than one abstractor; and the quality of the assessment of the

RCTs included in the MAs (16.8%).

Paper-based MAs versus Cochrane Publications

Items from the Oxman and Guyatt Scale (1991) revealed significant differences between these two types of MAs. These differences were not always statistically significant, but a number of them were greater than 10%, a level that, according to the consensus of experts canvassed, is substantively significant. A difference of -32.7 % was found in the reporting of whether appropriate search methods were used in the collection of evidence. Superior reporting in this area was more likely on paper-based than on electronically published MAs. A similar result was obtained when reviewing whether the search strategy used to collect evidence was reasonably comprehensive. In this case there was a 15.5% difference in favour of paper-published reports.

Paper-based MAs were also stronger in their reporting of the assessment of the validity of the RCTs they included. They were more likely to report the criteria used for assessing the validity of studies, and to use appropriate criteria to assess the validity of studies (either in selecting studies for inclusion or in analysing the studies cited). They were also more likely to report the methods used to combine the findings of the relevant studies to reach a conclusion. Many of the differences observed in this area were both clinically and statistically significant.

When assessing the quality of the reports of paper-published versus Cochrane MAs using the checklist of Sacks et al., the paper-published MAs obtained better scores. For example, in the reporting of the literature search, a statistically significant difference of 32.7% was obtained.

There was no difference in the results when the “partitioning the degrees of freedom” procedure was carried out. (See Tables 12, 13 and 14)

Table 12

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Oxman and Guyatt Overview Quality Assessment Questionnaire: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results

Question	Yes + Partially Or can't tell Paper-based	Yes + Partially or can't tell Cochrane	No Paper-based	No Cochrane	P-Value
1. Were the search methods used to find evidence reported?	41 [78.8]	51 [98.1]	11 [21.2]	1 [1.9]	.0022
2. Was the search strategy for evidence reasonable comprehensive?	33 [63.5]	49 [94.2]	19 [36.5]	3 [5.8]	.0001
3. Were the criteria used for deciding which studies to include in the overview reported?	45 [86.5]	52 [100.0]	7 [13.5]	0	.0062
4. Was bias in the selection of studies avoided?	38 [73.1]	38 [73.1]	14 [26.9]	14 [26.9]	1.000
5. Were the criteria used for assessing the validity of the included studies reported?	30 [57.7]	48 [92.3]	22 [42.3]	4 [7.7]	.0001
6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?	49 [94.2]	49 [94.2]	3 [5.8]	3 [5.8]	1.000
7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	49 [94.2]	41 [78.8]	3 [5.8]	11 [21.2]	.0215
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?	52 [100.0]	51 [98.1]	0	1 [1.9]	.3150
9. Were the conclusions made by the author (s) supported by the data and /or analysis reported in the overview?	51 [98.1]	52 [100.0]	1 [1.9]	0	.3150

*See Appendix C for a full description of each item

Table 13

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results

Questions	Adequate Later	Adequate Cochrane	Partial + No Later	Partial + No Cochrane	P-Value
1. Prospective Design					
a. Protocol	3 [5.8]	2 [3.8]	49 [94.2]	50 [96.2]	.6467
b. Literature search	33 [63.5]	16 [30.8]	19 [36.5]	36 [69.2]	.0008
c. Lists of trials analyzed	41 [78.8]	50 [96.2]	11 [21.2]	2 [3.8]	.0076
d. Log of rejected trials	17 [32.7]	31 [59.6]	35 [67.3]	21 [40.4]	.0059
e. Treatment assignment	33 [63.5]	49 [94.2]	19 [36.5]	3 [5.8]	.0001
f. Ranges of patients	19 [36.5]	17 [32.7]	33 [63.5]	35 [67.3]	.6802
g. Ranges of treatment	22 [61.5]	50 [96.2]	20 [38.5]	2 [3.8]	.0000
h. Ranges of diagnosis	22 [42.3]	38 [73.1]	30 [57.7]	14 [26.9]	.0015
2. Combinability					
a. Criteria	15 [28.8]	10 [19.2]	37 [71.2]	42 [80.8]	.2512
b. Measurement	34 [65.4]	15 [28.8]	18 [34.6]	37 [71.2]	.0002
3. Control of bias					
a. Selection bias	7 [13.5]	10 [19.2]	45 [86.5]	42 [80.8]	.4263
b. Data-extraction bias	14 [26.9]	16 [30.8]	38 [73.1]	36 [69.2]	.6651
c. Inter-observer agreement	7 [13.5]	1 [1.9]	45 [86.5]	51 [98.1]	.0273
d. Source of support	3 [5.8]	45 [86.5]	49 [94.2]	7 [13.5]	.0000
4. Statistical analysis					
a. Statistical methods	44 [84.6]	51 [98.1]	8 [15.4]	1 [1.9]	.0146
b. Statistical errors	6 [11.5]	2 [3.8]	46 [88.5]	50 [96.2]	.1410
c. Confidence intervals	41 [78.8]	48 [92.3]	11 [21.2]	4 [7.7]	.0507
d. Subgroup analysis	30 [57.7]	39 [75.0]	22 [42.3]	13 [25.0]	.0618
5. Sensitivity analysis					
a. Quality assessment	22 [42.3]	21 [40.4]	30 [57.7]	31 [59.6]	.8422
b. Varying methods	13 [25.0]	5 [9.6]	39 [75.0]	47 [90.4]	.0381
c. Publication bias	9 [17.3]	4 [7.7]	43 [82.7]	48 [92.3]	.1382
6. Application of results					
a. Caveats	36 [69.2]	27 [51.9]	16 [30.8]	25 [48.1]	.0709
b. Economic Impact	3 [5.8]	1 [1.9]	49 [94.2]	51 [98.1]	.3078
7. Language	2 [3.8]	2 [3.8]	50 [96.2]	50 [96.2]	1.000

*See appendix F and G for a full description of each item

Table 14

Quality of Reports of Meta-analyses of Randomized Trials, as Assessed on the Sacks Quality Assessment Checklist: Comparison of Paper-based Data Published Between 1990 and 1995 and Cochrane Data Published Between 1993 and 1996: Partitioning Degrees of Freedom Results

Question	Partial + Adequate Later	Partial + Adequate Cochrane	None or Unknown Later	None or Unknown Cochrane	P-Value
1. Prospective Design					
a. Protocol	10 [19.2]	5 [9.6]	42 [80.8]	47 [90.4]	.1629
b. Literature search	42 [80.8]	50 [96.2]	10 [19.2]	2 [3.8]	.0141
c. Lists of trials analyzed	47 [90.4]	50 [96.2]	5 [9.6]	2 [3.8]	.2404
d. Log of rejected trials	24 [46.2]	32 [61.9]	28 [53.8]	20 [38.5]	.1156
e. Treatment assignment	37 [71.2]	50 [96.2]	15 [28.8]	2 [3.8]	.0006
f. Ranges of patients	23 [44.2]	25 [48.1]	29 [55.8]	27 [51.9]	.6940
g. Ranges of treatment	39 [75.0]	50 [96.2]	13 [25.0]	2 [3.8]	.0021
h. Ranges of diagnosis	33 [63.5]	49 [94.3]	19 [36.5]	3 [5.8]	
2. Combinability					
a. Criteria	29 [55.8]	18 [34.6]	23 [44.2]	34 [65.4]	.0302
b. Measurement	41 [78.8]	18 [34.6]	11 [21.2]	34 [65.4]	.0000
3. Control of bias					
a. Selection bias	13 [25.0]	12 [23.1]	39 [75.0]	40 [76.9]	.8185
b. Data-extraction bias	21 [40.4]	17 [32.7]	31 [59.6]	35 [67.3]	.4153
c. Inter-observer agreement	12 [23.1]	9 [17.3]	40 [76.9]	43 [82.7]	.4637
d. Source of support	4 [7.7]	45 [86.5]	48 [92.3]	7 [13.5]	.0000
4. Statistical analysis					
a. Statistical methods	51 [98.1]	51 [98.1]	1 [1.9]	1 [1.9]	1.000
b. Statistical errors	10 [19.2]	3 [5.8]	42 [80.8]	49 [94.2]	.0379
c. Confidence intervals	42 [80.8]	48 [92.3]	10 [19.2]	4 [7.7]	.0848
d. Subgroup analysis	36 [69.2]	45 [86.5]	16 [30.8]	7 [13.5]	.0335
5. Sensitivity analysis					
a. Quality assessment	29 [55.8]	52 [100.0]	23 [44.2]	0	.0000
b. Varying methods	19 [36.5]	8 [15.4]	33 [63.5]	44 [84.6]	.0139
c. Publication bias	20 [38.5]	6 [11.5]	32 [61.5]	46 [88.5]	.0015
6. Application of results					
a. Caveats	49 [94.2]	50 [96.2]	3 [5.8]	2 [3.8]	.6467
b. Economic Impact	5 [9.6]	7 [13.5]	47 [90.4]	45 [86.5]	.5393
7. Language	4 [7.7]	3 [5.8]	48 [92.3]	49 [94.2]	.6955

*See appendix F and G for a full description of each item

4.0 DISCUSSION AND CONCLUSIONS

4.1 Major research objectives

4.1.1 Objective one - Quality of reporting on paper-based MAs (1977-1995)

Given the validity of the two assessment tools used, the quality of the reporting in the examined MAs was low. The results obtained in this study when assessing the group of MAs from the 'Later Years' using the Oxman and Guyatt Scale (1991) are slightly lower than those published by other authors. As previously noted, Jadad et al. (1996) found a median of four for the studies they evaluated in the area of pain. Sacks et al. (1987) reported slightly higher results for the same time periods. The results obtained varied slightly as a function of the particular MAs assessed.

The paper-based MAs assessed generally describe the methods they used to find evidence. What they often leave under-reported and unclear is the comprehensiveness of the search conducted. Comprehensiveness of search is included in the Oxman and Guyatt Scale, because its developers considered it to be an important issue in assessing the quality of an MA. Three quarters of the paper-based MAs stated that they had relied on more than just MEDLINE for their search.

The Sacks et al. (1987) checklist gives the issue of protocols considerable weight. Only a handful of the paper-based MAs reported that they had indeed developed an 'a priori' protocol. The issue of protocols is dealt with less stringently by the Oxman and Guyatt tool.

The MAs examined scored reasonably well on establishing 'a priori' criteria for the inclusion and exclusion of studies, although, the designing of an 'a priori' protocol was not clearly reported.

An area that was extremely under-reported was the control of bias in the selection of studies. A very low percentage of the MAs examined clearly stated that more than one reviewer had participated in the selection of the studies included in the review. In order to gain the

confidence of readers that bias has been avoided, it is generally recommended that a minimum of two reviewers apply inclusion and exclusion criteria, extract data and assess the quality of all included studies.

Many meta-analysts consider the assessment of the internal and external validity of studies to be a crucial step in the conduct of meta-analysis. They also believe that the value of such assessments depends upon the use of appropriate criteria.

Only half of the MAs assessed reported assessing the quality of their included studies, and the criteria used in their assessment. However, most of those that mentioned assessing the quality of studies also reported using appropriate criteria for this assessment.

How an MA handles the issue of combinability is important. Often reviewers struggle with the question of whether or not studies should be pooled. A very large percentage of the MAs assessed reported the criteria used for deciding whether studies were similar enough to pool.

A high proportion of MAs also reported the methods they used to combine the findings of individual studies. Those that did so were usually found to have combined these findings appropriately relative to the primary question.

A very low percentage of MAs reported on the blinding of the studies they surveyed. The issue of blinding remains a controversial area, with evidence being adduced both in favour of blinding and against it.

It is a legitimate concern that only a handful of the MAs assessed reported their sources of financial support, or clearly acknowledged potential conflicts of interest. This was somewhat surprising, given that journals generally encourage the inclusion of conflict of interests statements.

A small percentage of MAs reported conducting a sensitivity analysis, and an even smaller proportion reported using a method to calculate the number of unpublished negative studies that would be required to refute the published evidence.

Two other areas where MAs scored very poorly were the assessment of the economic impact of their findings, and the inclusion of studies in languages other than English. Until recently, these issues have not been viewed as particularly significant in the conduct of meta-analysis. The language item was assessed in this study because of the findings reported by Egger et al. (1997). This research suggested that the majority of “negative” trials seem to be published in non-English journals. The systematic exclusion of “negative” trials from inclusion in MAs on linguistic grounds may tend to bias an MA toward “positive” conclusions.

4.1.2 Objective two - Quality of reporting on paper-based MAs from the ‘Early Years’ versus the ‘Later Years’

Using the Oxman and Guyatt Scale (1991) the results reveal a trend, although not a statistically significant one, towards improvement in the overall quality of reporting of meta-analysis over time. However, the review of MAs conducted in this study indicates that the quality of their reporting remains at an unacceptably low level. And this is in spite of the substantial numerical increase in the production and publication of meta-analyses.

4.1.3 Objective three - Quality of reporting on paper-based versus electronically published MAs

In comparing Cochrane MAs to paper-based MAs, no significant difference was found in the overall quality of reporting. All MAs received extremely low scores for their reporting of items such as ‘a priori’ protocol, languages of included MAs and evaluation of economic impact. Less than four percent of all paper-based and Cochrane MAs, reported that an ‘a priori’ protocol had been developed. Before a completed Cochrane MA can be published, a peer-reviewed protocol must be published on The Cochrane Library. However, this protocol is usually not mentioned in published Cochrane MAs.

Cochrane reviews were better at reporting some items, and paper-based reviews at reporting others. Cochrane MAs were less likely to adequately report on their search methods and methods used to combine findings. On the other hand, Cochrane MAs were found to be more likely to report appropriately on the following issues: list of trials analyzed and log of

trials rejected; methods of treatment assignment; ranges of patients' characteristics; sources of support; and use of appropriate statistical methods and subgroup analysis. The assessment items on which reporting quality differed by more than 10% are listed above in section 3.3.5.

4.2 *Minor Research Objectives*

4.2.1 *Objective four – Development of effective search strategies for retrieving MAs*

Developing and running search strategies for this study provided valuable insights into how to design and refine an effective search strategy and use the information contained in abstracts to select studies for more detailed assessment. Many lessons were learned from the major difficulties encountered in distinguishing between a review, a literature review, an annotated bibliography, a systematic review and a meta-analysis. It became clear how important it is to choose words carefully and pay attention to detail when selecting titles for similar pieces of work. If the title and purpose of a project are not clearly stated by its author, it is very difficult for the reader to determine the author's intentions. This can lead to confusion, misinterpretation and the inappropriate application of findings.

4.2.2 *Objective five - Comparison of the findings of this study with findings reported in the published literature*

The development of many of the instruments created to assess the quality of MAs was found to be methodologically weak. The scale developed by Oxman and Guyatt (1991) is a notable exception to this observation. Its developers defined the construct they were interested in creating, measured the discriminatory power of items developed, and conducted inter-observer reliability studies as part of the development process. This scale also proved to be easy to use, and required only half as much time to apply as the other instrument used in this study. However, it is not the overall quality score yielded by a scale that most accurately assesses the validity of an MA, but the sum total of the information yielded by its individual components.

The extensive use made in this study of both a scale and a checklist, enabled raters to appreciate the benefits of each. Scores measuring the overall quality of MAs may be easy to compute and compare, but they don't give the various components. Useful insights and an awareness of potential difficulties are more likely to arise from the detailed examination of micro-level issues.

One of the biggest problems encountered in this study was the lack of published guidance on the application of the two quality assessment tools selected for use. After three rounds of pilot testing, and a few reviews of questions that arose concerning the application of the Oxman and Guyatt assessment tool, the assessors were ready to apply it. Although the Sacks checklist was very long and time consuming to apply, it was accompanied by more detailed directions as to its use.

4.3 *Study Limitations*

This study has a number of limitations that deserve mention. The study investigator and two content experts who contributed to the project are closely involved in the Cochrane Collaboration. Efforts were made to minimise the influence of personal bias when assessing the quality of MAs. These included the conduct of a pilot study, specifically designed to address this issue. A further limitation of the study is that The Cochrane Library was published relatively recently. Consequently, the randomly selected MAs assessed were likely to be from the same review groups and in some circumstances, by the same reviewer.

A pilot test was conducted with the assistance of an external reviewer. Three assessors participated in the application of the two tools used in this study to assess the quality of reporting of MAs. However, only one reviewer assessed all 151 of the MAs reviewed. External validity was assessed by a comparison of results obtained with the findings of similar previously published work.

A post hoc calculation was done to determine the adequacy of the size of the samples used in this project. A mean difference of .30 was detected between the paper-based 'Early Years' MAs and the paper-based 'Later Years' MAs, on the Oxman and Guyatt question assessing

overall quality (i.e. #10). A post hoc calculation revealed that one would need to include 565 MAs in each group in order to detect a statistically significant difference of this size by chance alone. Similarly, in comparing the paper-based 'Later Years' MAs to the Cochrane MAs, a mean difference of .07 was found. A post hoc calculation revealed that one would need to include 5081 MAs in each group in order to detect a statistically significant difference.

Another limitation, which should be addressed, is the time periods chosen for this study. It was unfortunate that equal numbers of MAs were not available from all three-time periods, but this was unavoidable. It was also unfortunate that the Cochrane MAs assessed were not published over exactly the same period of time as the paper-based MAs of the 'Later Years'. However, the time difference involved was minimal, and assumed unlikely to distort the results of this study.

A final limitation of this study was imposed by the decision to mask electronically-published (Cochrane) MAs in order to limit the influence of assessor bias. Electronically-published MAs had to be printed to permit their masking. This meant that the assessor could not access hyper-linked documents cited in these studies. Consequently, useful information relative to sources and the search strategies of the collaborative review group could not be considered in the assessment of the quality of reporting in these MAs.

4.4 *Implications of results*

The importance of the low scores obtained in the quality assessment of the MAs examined should not be minimized. These results indicate that the quality of reporting of MAs continues to be compromised by major flaws. This is a serious situation, because clinicians, health policy makers and consumers are repeatedly being told that MAs are, 'the best evidence available'. This may be true, but even if it is the best available, it is not good enough!

In spite of the criticisms which may be directed at meta-analyses, they have the potential to revolutionize the way literature reviews are conducted, by transforming the review process

from an art into a science (Cook et al., 1992; Rosenthal, 1991; Sipe, 1996).

Appropriately conducted and reported meta-analyses can help to resolve controversies between conflicting studies, guide clinical research by providing new hypotheses, identify areas in which insufficient research has been performed or in which additional research may be unethical, and identify beneficial and harmful therapies much earlier than other types of reviews (Antman et al., 1992).

Meta-analysis offers a possible way to keep up to date without sacrificing quality and thoroughness. Theoretically, it can effectively summarize the accumulated research on a topic, permit the consideration of fresh questions on the basis of available data, and channel the stream of clinical research towards relevant horizons. Consequently, meta-analyses are important to health policy planners and others involved in the provision of health care.

One of the advantages of meta-analysis touted by experts is that their replication is feasible, because their authors use specific protocols to search and gather studies. However, many details concerning search procedure were unreported in the meta-analyses assessed. This may be due to factors such as editorial decisions, the form in which authors elect to report their studies or the amount of information available in primary studies. Nevertheless, such limited reporting of procedural information may hamper replication of a particular meta-analysis.

Just as meta-analysts have found that the reporting of primary research sometimes does not lend itself to meta-analysis, so also the reporting of the results of meta-analysis can be unacceptable for application by medical professionals. Poor reporting by primary researchers can only exacerbate this weakness in the reporting of MA, while good reporting of the conduct of primary studies can facilitate the work of the meta-synthesist (Sipe 1996).

Clearly, improvements in the way meta-analysis is currently reported are needed for the development and reporting of clinical guidelines and health policy. The meta-analytical report, as a “stand alone” document, is all the health professional has to base decisions on. If it fails to provide an accurate, inclusive, and convincing account of the process followed,

then what good is it?

4.4.1 Solutions

One reason for the low quality of reporting observed may be a lack of guidance in the reporting of meta-analytical work. While the theoretical description of the methodology of meta-analysis may have improved, the quality of reporting in MAs has not kept pace. The weakness may be greatest in the areas of reading, understanding and following available methodological guidance. The impact of the busy schedules and heavy workloads of researchers should not be discounted; and neglect of methodological niceties may also play a part. Meta-analysts must ask themselves whether they are paying enough attention to detail, and utilizing available resources to provide the reader with, “the best available evidence”.

High quality training for researchers conducting meta-analysis should be made more readily available internationally. This can be accomplished through the provision of more courses on MA, more workshops on the use of appropriate statistical software, and more presentations and workshops on methodology for meta-analysts, reviewers and editors of MA. Initiatives should also be pursued to encourage more efficient collaboration with content experts and methodological specialists in the field (i.e., library scientists, statisticians, policy makers and journal editors).

There is a widespread perception that our editors are failing us, but substantial evidence is lacking to support this view. Many journal editors are responding to suggestions about how they can serve the needs of their readers more effectively (Smith 1994 and Haynes 1991). One way they can do this is by accepting the need to improve the scientific quality of the MAs they publish (Huth 1987). There is no doubt that editors are in a position to promote improvement in the quality of published reports of meta-analysis.

Another way to improve the quality of reporting of MAs is to develop guidelines for their reporting. These guidelines should be evidence-based (i.e., focused on the proper collection, management, interpretation and reporting of evidence) whenever possible, and aligned with leading-edge developments in the improved reporting of research. The Consolidation of

Structured Reporting of Randomized Trials (CONSORT) guidelines (Begg, 1996) developed, in part, by the Standards of Reporting Trials Group, represent a valuable step forward in the methodology of reporting on RCTs. What the CONSORT guidelines are designed to do for the reporting of RCT, the QUOROM statement has been developed to do for the reporting of meta-analysis.

The QUOROM (i.e., Quality of Reporting of Meta-analyses) statement is an evidence-based standard for the reporting of meta-analysis. It includes an eighteen-item checklist and flow chart that clearly and definitively describe the meta-analyst's task. The checklist includes eight evidence-based items that address the abstract, introduction, methods and results sections of a report of a meta-analysis of randomized trials. The methods and results sections provide information regarding searching, selection, quality assessment, data abstraction, qualitative and quantitative data synthesis, and trial flow.

The systems currently used for peer review may also benefit from fine-tuning to encourage improvement in the reporting of meta-analysis. The Cochrane Collaboration has begun to achieve this objective through a combination of continual peer review throughout the systematic review process, and the use of strict criteria that must be reflected in the meta-analytical process and included in the report. Paper-based MAs face additional obstacles. However, the use of evidence-based criteria (e.g., the QUOROM statement) may help to improve their quality as well.

4.5 *Significance of Research*

The evidence gathered by this project has made possible a fuller understanding of currently available tools for assessing the quality of reporting and general characteristics of MAs, and the limitations of these tools. This data has also provided helpful insights and knowledge about the reporting of meta-analysis.

Specifically, the results of this project provide:

- a) a clearer understanding of the quality of reporting of MAs;

- b) information on whether there has been improvement over time in the quality of reporting of MAs;
- c) information on the quality of paper-based MAs versus those published in The Cochrane Library; and
- d) empirical evidence on the kinds of tools (e.g., checklists and scales) that are useful for assessing the quality of reporting of MAs.

A 'key' test of meta-analyses and randomized controlled trials is whether their written reports persuade the reader to have confidence that the report is an accurate reflection of the methodological process followed during their various stages. Over the past decade, inroads have been made toward improving the quality of reporting of RCTs. The results of this study suggest that similar attention is required to improve the quality of reporting of meta-analyses (Moher 1995).

This study has also revealed that the quality of reporting of MAs has tended to improve slightly over time; and that the quality of the reporting of Cochrane MAs is no better or worse than that of paper-based reports, when measured by quality assessment tools. This is somewhat surprising, given that the avowed mission of the Cochrane Collaboration is to adduce, 'the best available evidence'.

It remains to consider how we can improve both the quality and the reporting of meta-analysis for the next generation of health-care decision-makers.

List of Meta-analyses

For a list of the Meta-analysis included in this study, see Appendix J.

References

- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA* 1992; 268:240-248.
- Assendelft WJJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA* 1995; 274:1942-1948.
- Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? *Stat Med* 1987; 6:351-358.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkkin R, Rennie D, Schultz Kf, Simel D, Stroup D. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996; 276:637-39.
- Berlin JA on behalf of the university of Pennsylvania meta-analysis blinding study group. Does blinding of reader's affect the results of meta-analyses? *Lancet* 1997; 350:185-186.
- Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1359-61.
- Cook DB, Sackett DL, Spitzer W. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol.* 1995; 48:167-171.
- Cook DJ, Mulrow C, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med*, 1997; 126:376-380.
- Chalmers I, Haynes RB. Reporting, updating, and correcting systematic reviews of the effects of health care. In: Chalmers I, Altman DG, eds. *Systematic reviews*. London: BMJ Publishing Group. 1995; 86-95.
- Dickersin, K., Berlin, J.A. (1992). Meta-analysis: State-of-the-Science. *Epidemiologic Reviews*, 1992; 14:154-176.
- Donner A, Eliasziw M. Sample Size Requirement for Reliability Studies. *Statistics in Medicine*, 1987; 6:441-448.
- Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Languages bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326-329.

Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? *Journal of Clinical Epidemiology* 1995; 48:159-163.

Guyatt GH, Sackett DL, Sinclair J, Hayward R, Cook DJ, Cook RJ. Users guide to the medical literature. IX. A method for grading health care recommendations. *JAMA*; 1995; 274:1800-1804.

Haynes RB. How clinical journals could serve clinician readers better. In: Lock S, ed. *The future of medical journals*, London:BMJ 1991:116-26.

Huth EJ. Needed: review articles with more scientific rigor. *Ann Intern Med* 1987; 104:485-8.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DMJ, et al. Assessing the quality of randomized clinical trials: is binding necessary? *Control Clin Trials* 1996;17:1-12.

Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol*; 1996;49:235-243

Jadad AR, Cook DJ, Jones AL, Klassen T, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: A comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280(3):278-280.

Klassen T, Jadad AR, Moher D. Guides for reading and interpreting systematic review: 1. Getting started. *Arch Ped & Adolescent Med* 1998; 152:700-704.

L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med*; 1987;107:224-233.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.

Light, RJ, Pillemer, DB *A Checklist For Evaluating Reviews*. In *The Science of Reviewing Research* 1984, (pp. 160- 186). Cambridge, Massachusetts and London, England: Harvard University Press.

Moher D, Berlin J. Improving the reporting of randomised controlled trials. In:Maynard A, Chalmers I, ed. *Non-random Reflections on Health Services Research*. London: BMJ 1997: 250-271.

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. Assessing the quality of randomized controlled trails: implications for the conduct of meta-analyses. *Health Tech Assess* 1999;3(4).

- Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Tech Assess Health care* 1996; 12:195-208.
- Moher D, Olkin I. Meta-analyses of randomized controlled trials: a concern for standards. *Journal of the American Medical Association*, 1995; 274:1962-1964.
- Mulrow CD. The Medical Review Article: State of the Science. *American College of Physicians*, 1987; 106:485-488.
- Oxman AD. Checklists for review articles. *British Medical Journal*, 1994; 309:648-651.
- Oxman AD, Cook DJ, Guyatt G. Users' Guides to the Medical Literature. *JAMA*, 1994; 272(17): 1367-1371.
- Oxman AD, Guyatt G. Guidelines for reading literature reviews. *CMAJ*, 1998; 138:697-703.
- Oxman AD, Guyatt G, Singer J, Goldsmith CH, Hutchison BG, Milner RA, Streiner DL. Agreement among Reviewers of Review Articles. *J Clin Epidemiol*, 1991; 44(1):91-98.
- Pignon JP, Arriagada R. Meta-analyses of randomized clinical trials: how to improve their quality? *Lung Cancer*, 1994; 1:S135-S141.
- Sacks H, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of Randomized Controlled Trials. *The New England Journal of Medicine*, 1987; 316(8): 450-455.
- Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. *Mount Sinai J Med*; 1996; 63:216-224.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical Evidence of Bias. *JAMA*, 1995; 273(5): 408-412.
- Shea B, Dube C, Moher D. Assessing the Quality of Reports of Systematic Reviews and Meta-analyses: A Systematic Review of Checklists and Scales. In Egger M, Altman DG. *Systematic reviews: 2nd edition*. BMJ Publishing group, 1999.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979; 86:420-428.
- Silagy CA. An Analysis of Review Articles Published in Primary Care Journals. *Family Practice-An International Journal*, 1993; 10(3):337-340.
- Sipe TA, Curlette WL. A Meta-synthesis of Factors Related to Educational Achievement: A Methodological Approach to Summarizing and Synthesizing Meta-analysis. *International Journal of Educational Research*, 1996; 25(7):583-698.

The Standards of Reporting Trials Group, A Proposal for Structured Reporting of Randomized Controlled Trials. JAMA; 1994; 272(24):1926-1931.

Siegel S, Castellan NJr. Nonparametric Statistics for the Behavioral Sciences. 2nd ed. Toronto: McGraw-Hill 1988.

Smith R. What's wrong with the BMJ? (editor's choice). BMJ 1994 January 8, 308:i.

Appendix A

Cochrane Collaboration Glossary

Bias

A systematic error (or bias) occurs when there is tendency to produce results that differ in a systematic manner from true values. The key term in understanding the concept of bias is 'different'. If, for example, the way in which participants are allocated to a treatment/intervention or control group is 'different', then there is a possibility that the observed effects of treatment are due to the incomparability of the two groups rather than to a true effect of the treatment/intervention. The implication is, therefore, that systematic errors may be responsible for the observed changes in outcome. There are many potential sources of bias, some of which can be controlled or prevented through careful study design. For example, the design of the randomized controlled trial reduces allocation bias. Participants in such a study are randomly allocated to one or more treatments/interventions, which maximize the probability that the groups receiving these different interventions will be comparable. See also methodological quality, validity.

Blinding (Syn. Masking)

In clinical trials this means that the participants, and often the investigators, are not aware which treatment/intervention the participant is getting. The purpose of blinding is to reduce ascertainment bias. For example, if the investigator knew which treatment/intervention a participant was getting, (s)he might intentionally or unintentionally respond differently to participants getting each treatment/intervention (e.g., in the recording of measurements). See also single blind, double blind and triple blind.

CDSR (Cochrane Database of Systematic Reviews)

One of the products of the Cochrane Collaboration. It brings together all the currently available Cochrane reviews and is available on CD-ROM and the internet. It also contains general information about the Cochrane Collaboration, and more specific details of Methods Groups and Collaborative Review Group. It is issued quarterly. Collaborative Review Groups submit modules of edited reviews and other information to the Parent Database for inclusion in the CDSR.

CINAHL (Cumulative Index of Nursing and Allied Health Literature)

An electronic database covering all the major journals in nursing and allied health. Virtually all relevant English language publications are indexed. Years of coverage: 1983-present.

Clinical trial (Syn. therapeutic trial, interventional study)

A trial which tests out a drug or treatment/intervention to see whether it is effective or safe. This general term encompasses randomized controlled trials, controlled clinical trials, and randomized clinical trials.

Cochrane Collaboration

An international endeavour in which people from many different countries systematically find, appraise and review available evidence from RCTs and other sources of reliable evidence. The Collaboration aims to develop and maintain systematic, up-to-date reviews of RCTs of all forms of health care, and to make this information readily available to clinicians, consumers and other decision-makers at all levels of health care systems.

Cochrane Collaboration Handbook

The main working document of the Cochrane Collaboration. Its contents reflect decisions made at the annual Colloquia and by the Steering Group. The Handbook consists of the following six sections:

Section I	Background, aims and organization of the Cochrane Collaboration
Section II	Establishing and supporting Collaborative Review Groups
Section III	Representing the interests of Fields
Section IV	Cochrane Centers
Section V	Establishing and maintaining registers of RCTs
Section VI	Preparing and maintaining systematic reviews (see Tool Kit)

The Handbook can be down loaded from FTP servers both at the Canadian Cochrane Center and the UK Cochrane Center.

Cochrane Database of Systematic Reviews (CDSR)

See CDSR.

Concealment of Allocation

The process used to prevent foreknowledge of treatment/intervention assignment. The process should be impervious to any influence by the individual making the allocation. Most likely to be achieved securely if the randomization process is administered by someone who is not responsible for recruiting participants, for example, a hospital pharmacy, or a central office. Methods of assignment such as date of birth and case record numbers (see quasi random) are open to manipulation by the trial investigator.

Confidence Interval (CI)

The range within which the true size of effect of a treatment or intervention (never exactly known) lies with a given degree of assurance. People often speak of a '95% confidence interval' (or '95% confidence limits'). This is the interval which includes the true value in 95% of cases.

Current Contents

Electronic database which provides access to the tables of contents and bibliographic data of current issues of the world's leading scholarly research journals in the sciences, social sciences, arts and humanities. Over 6,600 journals covered.

Double Blind (Syn. Double Masked)

Neither the participant in a trial nor the investigator assessing the patient outcomes is aware of the treatment/intervention the participant is being given. Double blind is also used to describe trials where, in addition, the investigators carrying out the treatment/intervention are unaware of whether they are providing experimental or control treatments/interventions. The purpose of this is to reduce bias. A double blind trial is usually a randomized controlled trial or a controlled clinical trial. See also blinding, single blind, triple blind.

EMBASE (Excerpta Medica database)

A European electronic database of pharmacological and biomedical literature covering 3,500 journals from 110 countries. Years of coverage - 1974 to present.

Handsearching

Handsearching within the Cochrane Collaboration refers to the systematic searching of a journal page by page, including editorials, letters, etc., to identify all reports of randomized controlled trials. Normally a person would start handsearching the journal with the current year, and work backwards to 1948 (or volume 1 if after 1948). Once a trial is found, it is coded appropriately using definitions agreed upon within the Cochrane Collaboration, and outlined in Section V of the Cochrane Collaboration Handbook. The results of the search are sent to the Baltimore Cochrane Center, which forwards them to the National Library of Medicine for re-tagging on MEDLINE. A handsearching manual is available through the Baltimore Cochrane Center, and should be read together with Section V of the Cochrane Collaboration Handbook before handsearching is commenced. A journal handsearching registration form must be completed for each journal title, and sent to Baltimore to avoid duplication of effort.

Index Medicus

Catalogue of the United States National Library of Medicine (NLM), and a periodical index to the current literature. Available in paper form, or electronically as part of MEDLINE.

MEDLINE (MEDLARS onLINE)

An electronic database produced by the United States National Library of Medicine, which summarizes millions of pieces of biomedical research literature, in selected journals (contains articles from about 3,700 journals). It is available through most health service libraries. It can be accessed on CD-ROM, the Internet and by other means. Years of coverage - 1966 to present.

MeSH Headings (=Medical Subject Headings)

Terms used by the United States National Library of Medicine to index articles in Index Medicus and MEDLINE. Designed to reduce problems that arise from, for example, differences in British and American spelling. The MeSH system has a tree structure in which broad subject terms branch into a series of progressively narrower subject terms.

Meta-analysis

A statistical technique which summarizes the results of several studies into a single estimate, giving more weight to results from larger studies.

Methodological Quality

The extent to which the design and conduct of a trial are likely to have prevented systematic errors (bias). Variation in quality can explain variation in the results of trials included in a systematic review. More rigorously designed (better 'quality') trials are more likely to yield results that are closer to the 'truth'. While many criteria are used to judge methodological quality, only one of these, concealment of allocation, is presently based on strong empirical evidence. See also external validity, validity, study validity, bias.

Quality Assessment

Quality assessment of the controlled trials being considered for inclusion in systematic reviews is needed to maximize the validity of reviews. Although trial quality can be assessed from on or more of a variety of perspectives, the aspect which has been studied most thoroughly is the extent to which biases have been avoided in the comparisons made with the trial. Trial quality assessment is a focus of one of the Cochrane Methods Working Groups.

Randomization

Method used to generate the random allocation sequence. Includes tables of random numbers, computer-generated numbers, etc. The method of randomization is distinguished from random allocation. For instance, a list of random numbers may be used to generate a random sequence of treatments/interventions that are in turn allocated to participants, using sealed envelopes.

Relative Risk (RR)

One way of describing the effect of treatment/intervention where the outcomes are events. It is the ratio of the risk (or rate) of the event in the treatment group to the risk (or rate) of the event in the control group. Relative risks above (below) 1 indicate that the event is more (less) likely to occur in the treatment group than in the control group. When the event rate is small relative risks are very similar to odds ratios, and then the terms are often used interchangeably.

Review

Has various meanings:

Systematic review.

- A. In the medical literature, this is an article, which looks at a number of different studies and may draw conclusions about a particular treatment or intervention. The review may or may not be looking at randomized controlled trials. This type of review is often not systematic.
- B. To referee a paper. See referee process, external peer reviewer.

Search Activity

The activity undertaken by a Collaborative Review Group to identify the greatest number of trials. This may include handsearching relevant journals, searching electronic databases, contacting drug companies, cross-referencing journal articles, etc. Collaborative Review Groups must describe their search activity in detail in ModMan. Reviewers can then refer to the Group's search activity when writing a review, and if necessary supplement it with their own additional search. See also search strategy.

Search Strategy

Strategy used to identify and retrieve citations of reports of randomized controlled trials from electronic databases. A detailed search strategy to retrieve such reports from MEDLINE has been devised by Ms Carol Lefebvre and colleagues at the UK Cochrane Center, and can be found in the Cochrane Handbook.

Statistical Power

Statistical power is a measure of a study's ability to detect a clinically important difference based on the number of participants who are entered.

Statistically Significant

The findings of a study may be just an unusual fluke. The P-value from statistical tests tells us how likely it is that the results are just chance findings. Significance is merely the level at which we decide to take something seriously (1 in 20=5%, 1 IN 100=1%).

Study Validity (Syn. Internal Validity)

The degree to which the results of a study are likely to be true. Study validity depends on methodological quality. Randomized controlled trials which ensure allocation concealment, blinded assessment and relatively complete follow-up usually have high study validity.

Systematic Review (Syn. Systematic Overview)

In the Cochrane Collaboration, it is a review in which evidence (usually from randomized controlled trials) on a topic has been systematically identified, appraised and summarized according to predetermined criteria. Such reviews can be systematic (taking steps to reduce bias) without using statistical synthesis (meta-analysis) to reduce imprecision.

Validity

Validity is the degree to which any result (of a measurement or study) is likely to be true and free of systematic bias. See also study validity, external validity, methodological validity.

Appendix B

List of Checklists and Scales

1. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC: Meta-analyses of randomized controlled trials. *The New England Journal of Medicine* 1987;316:450-454.
2. Oxman Ad Guyatt GH. Index of the scientific quality of research overviews. Unpublished.
3. Meyer TJ, Mark MM. Effects of psychosocial interventions with adult cancer patients: A meta-analysis of randomized experiments. *Health Psychology* 1995;14:101-108.
4. Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *British Medical Journal*. 1997;315:149-153.
5. Salomon P, Kornbluth A, Aisenberg J, Janowitz HD. How effective are current drugs for Crohn's disease? A meta-analysis. *Journal of Clinical Gastroenterology*. 1992;14:211-215.
6. Qizilbash N, Birks J, Arrieta JL, Lewington S, Szeto S. The efficacy of tacrine in Alzheimer's disease. *Cochrane Database of Systematic Reviews*
7. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *International J Epidemiology*. 1999;28:1-9.
8. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epid* 1995;48:167-171.
9. Geller NL, Proschan M. Meta-analysis of clinical trials: a consumer's guide. *J. Biopharm Stat*. 1996;6:377-394.
10. Goldschmidt PG: Information Synthesis: A practical guide. *Health Services Research* 1986;21:215-237.
11. Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). In: *How to read a paper – the basics of evidence based medicine*. BMJ Publishing Group, London. 1997.
12. Taylor Halvorsen K. The reporting format. edited by Cooper H and Hedges LV, 1994, p. 425-437.
13. Irwig L, Tosteson ANA, Gatsonis C, et al: Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine* 1994;120:667-676.

14. Light RJ, Pillemer DB: The Science of reviewing research. Harvard University Press 1984;160-186.
15. Mullen PD, Ramirez G: Information synthesis and meta-analysis. *Advances in Health Education and Promotion* 1987;2:201-239.
16. Mulrow CD: The medical review article: state of the science. *Annals of Internal Medicine* 1987;106:485-488.
17. Neely JG: Literature review articles as a research form. *Otolaryngology - Head & Neck Surgery* 1993;108:743-748.
18. Nony P, Cucherat M, Haugh MC, Boissel JP: Critical reading of the meta-analysis of clinical trials. *Therapie* 1995;50:339-351.
19. Ohlsson A: Systematic reviews - theory and practice. *Scand J Clin Lab Invest* 1994;54:25-32.
20. Oxman AD: Checklists for reviews articles. *British Medical Journal* 1994;309:648-651.
21. Oxman AD, Guyatt G: Guidelines for reading literature reviews. *Canadian Medical Association Journal* 1988;138:697-703.
22. Pogue JM and Yusuf S. overcoming the limitations of current meta-analysis of randomized controlled trials. *The Lancet* 1998;351:47-52.
23. Smith MC, Stullenbarger E. Meta-analysis: an overview. *Nursing Science Quarterly*. 1989;2:114-115.
24. Thacker SB, Peterson HB, Stroup DF: Meta-analysis for the obstetrician-gynecologist. *American Journal Obstetrics Gynecology* 1996;174:1403-1407.
25. Wilson A, Henry DA: Meta-analysis. *Med J Aust* 1992;156:173-187.
26. Assendelft WJJ, Koes B, Knipschild PG, Bouter LM: The relationship between methodolgical quality and conculsuions in reviews of spinal manipulation. *Journal of the American Medical Association* 1995;274:1942-1948.
27. Auperin A, Pignon JP, Poynard T: Critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. *Almeth Pharmacol Therapy* 1999;11:215-225.

Appendix C

Oxman and Guyatt's Scale - Overview Quality Assessment Questionnaire (OQAQ)

The purpose of this index is to evaluate the scientific quality (i.e., adherence to scientific principles) of research overviews (review articles) published in the medical literature. It is not intended to measure literary quality, importance, relevance, originality, or other attributes of overviews.

The index is for assessing overviews of primary ("original") research on pragmatic questions regarding causation, diagnosis, prognosis, therapy or prevention. A research overview is a survey of research. The same principles that apply to epidemiologic surveys apply to overviews: a question must be clearly specified, a target population identified and accessed, appropriate information obtained from that population in an unbiased fashion, and conclusions derived, sometimes with the help of formal statistical analysis, as is done in "meta-analysis". The fundamental difference between overviews and epidemiologic surveys is the unit of analysis, not the scientific issues that the questions in this index address.

Since most published overviews do not include a methods section it is difficult to answer some of the questions in the index. Base your answers, as much as possible, on the information provided in the overview. If the methods that were used are reported incompletely relative to a specific item, score that item as "partially". Similarly, if there is no information provided regarding what was done relative to a particular question, score it as "can't tell", unless there is information in the overview to suggest either that the criterion was or was not met.

1. Were the search methods used to find evidence (original research) on the primary question(s) stated?

☐ yes ☐ partially ☐ no

2. Was the search for evidence reasonably comprehensive?

☐ yes ☐ can't tell ☐ no

3. Were the criteria (inclusion/exclusion) used for deciding which studies to include in the overview reported?

☐ yes ☐ partially ☐ no

4. Was bias in the selection of studies avoided?

☐ yes ☐ can't tell ☐ no

5. Were the criteria (methodological quality) used for assessing the validity of the included studies reported?

☐ yes ☐ partially ☐ no

6. Was the validity of all studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?

☐ yes ☐ can't tell ☐ no

7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?

☐ yes ☐ partially ☐ no

8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?

☐ yes ☐ can't tell ☐ no

For question 8, if no attempt was made to combine findings, and no statement is made regarding the inappropriateness of combining findings, check "no". If a summary (general) estimate is given anywhere in the abstract, the discussion or the summary section of the paper, and it is not reported how the estimate was derived, mark "no" even if there is a statement regarding the limitations of combining the findings of the studies reviewed. If in doubt mark "can't tell".

9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?

☐ yes ☐ partially ☐ no

For an overview to be scored as "yes" on question 9, data (not just citations) must be reported that supports the main conclusions regarding the primary question(s) that the overview addresses.

10. How would you rate the scientific quality of the overview?

Extensive
flaws

☐

1

Major
flaws

☐

3

Minor
flaws

☐

5

☐

6

Minimal
flaws

☐

7

The score for question 10, the overall scientific quality, should be based on your answers to the first nine questions. The following guidelines can be used to assist with deriving a summary score.

If the "can't tell" option is used one or more times in the preceding questions, a review is likely to have minor flaws at best and it is difficult to rule out major flaws (i.e., a score of 4 or lower). If the "no" option is used on question 2, 4, 6 or 8, the review is likely to have major flaws (i.e., a score of 3 or less, depending on the number and degree of the flaws).

Appendix D

Sacks Quality Assessment Checklist for Meta-analyses

	Adequate	Partial Number (percent)	None or Unknown
1. Prospective design			
a) Protocol			
b) Literature search			
c) List of trials analyzed			
d) Log of rejected trials			
e) Treatment assignment			
f) Ranges of patients			
g) Ranges of treatment			
h) Ranges of diagnosis			
2. Combinability			
a) Criteria			
b) Measurement			
3. Control of bias			
a) Selection bias			
b) Data-extraction bias			
c) Inter-observer agreement			
d) Source of Support			
4. Statistical analysis			
a) Statistical methods			
b) Statistical errors			
c) Confidence intervals			
d) Subgroup analyses			
5. Sensitivity analysis			
a) Quality assessment			
b) Varying methods			
c) Publication bias			
6. Application of results			
a) Caveats			
b) Economic impact			
7. Language			

Appendix E

SET SEARCH

001 meta-anlysis.pt,sh.	2866
002 (meta-anal: or metaanal:).tw.	2296
003 (quantitativ: review: or quantitativ: overview).tw.	79
004 (systematic: review: or systematic: overview:).tw.	287
005 (methodologic: review: or methodologic: overview:).tw.	76
006 review.pt,sh. Or review:.tw. Or overview:.tw.	635672
007 (integrative research review: or research intergration:).tw.	30
008 7 or quantitativ: synthes:.tw.	61
009 1 or 2 or 3 or 4 or 5 or 8	3920
010 (medline or medlars).ti,sh,ab. or embase.tw.	2371
011 (scisearch or psycinfo or psychinfo).tw.	24
012 (hand search: or manual search:).tw.	169
013 (electronic database: or bibliographic database:).tw.	88
014 (pooling or pooled analys: or mantel haenszel).tw.	1986
015 (peto or der simonian or dersimonian or fixed effect:).tw.	259
016 (psychlit or psychlit).tw.	15
017 010 or 11 or 12 or 13 or 14 or 15 or	4663
018 6 and 17	1447
019 18 or 9	5109
020 random:.tw,sh,pt. or placebo:.tw,sh	175571
021 clinical trial).pt.	151285
022 randomized controlled trial.pt	57462
023 double-blind:.tw,sh	52249
024 20 or 21 or 22 or 23	248821
025 19 and 24	1467

Appendix F

Enhanced Version of Oxman and Guyatt's Scale (OQAQ)

Oxman and Guyatt's index of the scientific quality of research overviews.

The purpose of this index is to evaluate the scientific quality (i.e., adherence to scientific principles) of research overviews (review articles) published in the medical literature. It is not intended to measure literary quality, importance, relevance, originality, or other attributes of overviews.

The index is for assessing overviews of primary ("original") research on pragmatic questions regarding causation, diagnosis, prognosis, therapy or prevention. A research overview is a survey of research. The same principles that apply to epidemiologic survey apply to overviews: a question must be clearly specified, a target population identified and accessed, appropriate information obtained from that population in an unbiased fashion, and conclusions derived, sometimes with the help of formal statistical analysis, as is done in "meta-analyses". The fundamental difference between overviews and epidemiologic surveys is the unit of analysis, not the scientific issues that the questions in this index address.

Since most published overviews do not include a methods section it is difficult to answer some of the questions in the index. Base your answers, as much as possible, on the information provided in the overview. If the methods that were used are reported incompletely relative to a specific item, score that item as "partially". Similarly, if there is no information provided regarding what was done relative to a particular question, score it as "can't tell", unless there is information in the overview to suggest either that the criterion was or was not met.

Were the search methods used to find evidence (original research) on the primary question(s) stated?

yes partially no

Yes is given to meta-analysis reporting categories of sources, including years (e.g., databases-medline) used, and whether these categories were named (e.g., medline). Partial points are given for the category of sources (e.g., electronic, hand, register) are named.

Was the search for evidence reasonably comprehensive?

yes can't tell no

Yes is given if at least three categories, one of which must be electronic with key words stated, and any two others (e.g., hand, register) are reported. Key words and/or MESH terms must be stated.

Were the criteria (inclusion/exclusion) used for deciding which studies to include in the overview reported?

yes partially no

This item was thought to be reasonably explicit. If 2 or more items mentioned, yes, if <2 mentioned, partially, if none mentioned, no.

Was bias in the selection of studies avoided?

yes can't tell no

Yes is given if at least two reviewers independently assess for inclusion. A consensus must be reached.

Were the criteria (methodological quality) used for assessing the validity of the included studies reported?

yes partially no

It was felt that the issues relating to publication bias should not be included in the assessment of this. Yes is given to those meta-analysis reporting 'a priori' methods of validity assessment (eg., if the author(s) chose to include only randomized, double-blind, placebo controlled trials, or allocation concealment as inclusion criteria).

Was the validity of all studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?

yes can't tell no

This item relates to validity assessment. Yes is given if there is a description of any criteria (either internal or external) used either for inclusion, or for analysis (eg., sensitivity analysis).

Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?

yes partially no

This item was thought to be reasonably explicit.

Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?

yes can't tell no

For question 8, if no attempt was made to combine findings, and no statement is made regarding the inappropriateness of combining findings, check "no". If a summary (general) estimate is given anywhere in the abstract, or the summary section of the paper, and it is not reported how the estimate was derived, mark "no" even if there is a statement regarding the limitations of combining the findings of the studies reviewed. If in doubt mark "can't tell".

Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?

yes partially no

For an overview to be scored as “yes” on question 9, data (not just citations) must be reported that supports the main conclusions regarding the primary question(s) that the overview addresses. If the overview concerns diagnostic/prognostic tests, ‘retest is not required’ (this ensures that diagnostic/prognostic papers are not scored more rigorously than clinical papers).

How would you rate the scientific quality of the overview?

Extensive
flaws

☐

1

Major
flaws

☐

3

Minor
flaws

☐

5

Minimal
flaws

☐

7

The score for question 10, the overall scientific quality, should be based on your answers to the first nine questions. The following guidelines can be used to assist with deriving a summary score. If the “can’t tell” option is used one or more times on the preceding questions, a review is likely to have minor flaws at best and it is difficult to rule out major flaws (i.e., a score of 4 or lower). If the “no” option is used on question 2, 4, 6 or 8, the review is likely to have major flaws (i.e., a score of 3 or less, depending on the number and degree of the flaws).

Appendix G

Enhanced Version of Sacks Quality Assessment Checklist for Meta-analysis

<u>Quality Assessment Form</u>

MA# _____

Demographics:

MA Subject: Causation, Diagnosis, Prognosis, Therapy or Prevention

Title: _____

Authors: _____

Journal: _____

Year of Publication: _____

Language of Publication: _____

Number of included subjects: _____

Number of included studies: _____

“Assessing the Quality of Reporting of Meta-analysis of Randomized Control Trials”

Were the following items reported?

1. Prospective Design

a) Protocol	Adequate	Partial	None or unknown
b) Literature Search	Adequate	Partial	None or unknown
c) List of Trials Analysed	Adequate	Partial	None or unknown
d) Log of rejected Trials	Adequate	Partial	None or unknown
e) Treatment Assignment	Adequate	Partial	None or unknown
f) Ranges of Patients	Adequate	Partial	None or unknown
g) Range of Treatment	Adequate	Partial	None or unknown
h) Ranges of Diagnosis	Adequate	Partial	None or unknown

2. Combinability

a) Criteria	Adequate	Partial	None or unknown
b) Measurement	Adequate	Partial	None or unknown

3. Control of Bias

a) Selection Bias	Adequate	Partial	None or unknown
b) Data-Extraction Bias	Adequate	Partial	None or unknown
c) Inter-observer Agreement	Adequate	Partial	None or unknown

d) Sources of Support	Adequate	Partial	None or unknown
-----------------------	----------	---------	-----------------

4. Statistical Analysis

a) Statistical Methods	Adequate	Partial	None or unknown
b) Statistical errors	Adequate	Partial	None or unknown
c) Confidence Intervals	Adequate	Partial	None or unknown
d) Subgroup Analysis	Adequate	Partial	None or unknown

5. Sensitivity of Applicability

a) Quality Assessment	Adequate	Partial	None or unknown
b) Varying Methods	Adequate	Partial	None or unknown
c) Publication Bias	Adequate	Partial	None or unknown

6. Wide Applicability

a) Caveats	Adequate	Partial	None or unknown
b) Economic Impact	Adequate	Partial	None or unknown

7. Have the authors included studies in all languages in their review?

Adequate	Partial	None or unknown
----------	---------	-----------------

Appendix H

Guidelines for completing Sacks Quality Assessment Form

Study Design

Protocol

The questions to be answered, the criteria for inclusion in the study and the methods to be used should be established before hand. This must be apparent to the reader and not simply assumed.

Literature Search

Details of the search strategy should be provided and it is insufficient for the author to rely solely on computer searches of the literature. A computer search must have been supplemented by consulting current contents, reviews, textbooks, or experts in the particular field of study, and by reviewing the references in the trials found.

List of Trials analysed and log of Rejected Trials

The report of the meta-analysis should provide a list of the trials analysed, as well as a list of the trials excluded and the reasons for exclusion.

Treatment Assignment

The method of treatment assignment in the primary study must be specified.

Ranges of Patient Characteristics, Diagnosis, and Treatment

Data should be provided on the patients, diagnosis, therapies, and end points in the original studies. The ranges of patient characteristics in all the trials analysed (e.g. age, sex, relevant socio-economic data, or other diseases) should be included.

Combinability

Criteria

A statement of the criteria used for deciding whether or not the trials were similar enough to be pooled. The author should note any differences in the primary study and discuss how these differences affect the conclusions.

Measurement

Assessment of the use of a test for homogeneity.

Control and Measurement of Potential Bias.

Selection Bias

The decision to include a paper should be made by looking only at its methods and not its results, or by looking at the two separately under coded conditions.

Data-Extraction Bias

More than one observer extracted the data, each of whom was blinded to the various treatment groups through a coded photocopying process and then an inter-observer agreement was measured.

Source of Support

The source of support should be stated and potential conflict of interest clearly acknowledged.

Statistical Analysis

Statistical Methods

Adequate was acknowledged as meaning any recognized method of pooling the simple addition of successes across all trials to give an overall average, which was rated as partial.

Statistical Errors

An awareness of potential problems of type 1 or type 11 error.

Confidence Intervals

Authors must have included confidence intervals with the significant values.

Subgroup Analyses

To increase statistical power.

Sensitivity Analysis

Whether the differences were assessed using quality, subgroups of patients, etc.

Quality Assessment

The methodologic rigor and scientific quality of the papers should be assessed and considered in formulating recommendations.

Varying Methods

Each meta-analysis should include sensitivity analysis data that show how the results vary with the use of different assumptions, tests and criteria.

Publication Bias

A simple method should be carried out for calculating the number of unpublished negative studies required to refute the published evidence.

Application of Results

Caveats

The MA should attempt to put the results into perspective on the basis of all the considerations listed above.

Economic Impact

It is important to consider the economic impact of adopting new methods of treatment or diagnosis.

Appendix I

Logistic Regression (Sacks Question 1)

Total number of cases: 151 (Unweighted)

Number of selected cases: 151

Number of unselected cases: 0

Parameter

	Value	Freq	Coding	
		(1)	(2)	
YEARCUT				
90 - 95	.00	52	-.333	-.333
77 - 89	1.00	47	.667	-.333
Cochrane 96	2.00	52	-.333	.667

STUDIES

Number of studies <= 9	1	82	-.500
Number of studies > 9	2	69	.500

Dependent Variable. A1A

Beginning Block Number 0. Initial Log Likelihood Function

-2 Log Likelihood 78.804936

* Constant is included in the model.

Beginning Block Number 1. Method: Enter

Variable(s) Entered on Step Number

1. YEARCUT Year of Publication
2. STUDIES Number of Studies

Estimation terminated at iteration number 5 because
Log Likelihood decreased by less than .01 percent.

-2 Log Likelihood	75.793
Goodness of Fit	150.840
Cox & Snell - R ²	.020
Nagelkerke - R ²	.049

	Chi-Square	df	Significance
Model	3.012	3	.3898
Block	3.012	3	.3898
Step	3.012	3	.3898

----- Hosmer and Lemeshow Goodness-of-Fit Test-----

A1A = "no or partial" A1A = yes

Group	Observed	Expected	Observed	Expected	Total
1	42.000	41.349	1.000	1.651	43.000
2	8.000	8.651	1.000	.349	9.000
3	22.000	21.681	1.000	1.319	23.000
4	27.000	27.319	2.000	1.681	29.000
5	13.000	13.970	3.000	2.030	16.000
6	28.000	27.030	3.000	3.970	31.000

Chi-Square df Significance

Goodness-of-fit test 2.4784 4 .6485

Classification Table for A1A

The Cut Value is .50

	Predicted						
	"no or partial"		yes		Percent Correct		
	"	I	y				
Observed	+	-----	+	-----	+		
"no or partial"	"	I	140	I	0	I	100.00%
	+	-----	+	-----	+		
yes	y	I	11	I	0	I	.00%
	+	-----	+	-----	+		
				Overall			92.72%

----- Variables in the Equation -----

Variable	B	S.E.	Wald	df	Sig	R
YEARCUT			2.6080	2	.2714	.0000
YEARCUT (1)	.8703	.7413	1.3781	1	.2404	.0000

YEARCUT (2)	-.4215	.9711	.1884	1	.6643	.0000
STUDIES (1)	.0109	.6827	.0003	1	.9873	.0000
Constant	-2.6442	.3448	58.8203	1	.0000	

95% CI for Exp (B)

Variable	Exp (B)	Lower	Upper
YEARCUT (1)	2.3876	.5584	10.2092
YEARCUT (2)	.6561	.0978	4.4010
STUDIES (1)	1.0109	.2652	3.8531

Appendix J

List of Meta-analyses Included in this Study

1. Abramson MJ, Puy RM, Weiner JM. Is Allergen Immunotherapy Effective in Asthma? *American Journal of Respiratory and Critical Care Medicine* 1995;151(4):969-74.
2. Anderson IM, Tomenson BM. Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. *BMJ* 1995; 310(6992):1433-8.
3. Anonymous. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. Early Breast Cancer Trialists' Collaborative Group. *NEJM* 1989;319(26):1681-92.
4. Anonymous. Maximum androgen blockade in advanced prostate cancer: and overview of 22 randomised trials with 3283 deaths in 5710 patients. *Lancet* 1995;346(8970):265-9.
5. Anonymous. Effect of Breast-Feeding on Oral Rhesus Rotavirus Vaccine Seroconversion: A Meta-analysis. 1990.
6. Antcak-Bouckoms A, Joshipura K, Burdick E, Tulloch JFC. Meta-analysis of surgical versus non-surgical methods of treatment for periodontal disease. *J Clin Periodontol* 1993;20:259-268.
7. Appel LJ, Miller ER, Seidler AJ, Whelton PK. Does Supplementation of Diet with 'Fish Oil' Reduce Blood Pressure. *Arch Intern Med* 1993;153:1429-1438.
8. a'Rogvi-Hansen B, Boysen G. Glycerol in acute stroke. *The Cochrane Library* 1996;Issue 3.
9. Bernard-Bonnin AA, StachenkoS, Bonin D, Charette C, Rousseau E. Self-management teaching programs and morbidity of pediatric asthma: A meta-analysis. *Journal of Allergy and Clinical Immunology* 1995;95(1Pt 1):34-41.
10. Bobbio M. Does most myocardial infarction rehabilitation prolong survival? A meta-analytic survey. *Giornale Italiano di Cardiologia* 1989;19(11):1059-67.

11. Broussel R. Prophylaxis of upper gastrointestinal bleeding in intensive care units: A meta-analysis. *Annales de Radiologie* 1989;32(4Pt2):329-32.
12. Brown KH, Peerson JM, Fontaine O. Use of Nonhuman Milks in the Dietary Management of Young Children With Acute Diarrhea: A Meta-analysis of Clinical Trials. *Pediatrics* 1994;932(1):17-27.
13. Buring JE, Evans DA, Mayrent SL, Rosner B, Colton T, Hennekens CH. Randomized trials of aminoglycoside antibiotics: Quantitative overview. *Reviews of Infectious Diseases* 1988;10(5):951-7.
14. Buyse M, Zeleniuch-Jacquotte A, Chalmers TC. Chimiotherapie Adjuvante du Cancer Colorectal: Qu'Apprendre D'une Meta-analysis? *Bull. Cancer* 1989;76:1021-1028.
15. Byrne MM. Meta-analysis of early phase II studies with paroxetine in hospitalized depressed patients. *Acta Psychiatr. Scand* 1989;80(Suppl350):138.
16. Cales P. La place des B-bloquants dans le traitement preventif de la rupture de varices oesophagiennes. *Acta Gastroenterologica Belgica* 1990;53(4):409-17.
17. Cappelleri JC, Fiore LD, Brophy MT, Deykin D, Lau J. Efficacy and safety of combined anticoagulant and antiplatelet therapy versus anticoagulant monotherapy after mechanical heart-valve replacement: A meta-analysis. *American Heart Journal* 1995;130(3 pt.1):547.
18. Cappuccio FP, Siani A, Strazzullo P. Oral calcium supplementation and blood pressure: An overview of randomized controlled trials. *Journal of Hypertension* 1989;7:941-946.
19. Cates CJ. Chambers/nebulisers-acute asthma. *The Cochrane Library* 1996;Issue 3.
20. Chalmers TC, Matta RJ, Smith H, Kunzler AM. Evidence favouring the use of anticoagulants in the hospital phase of acute myocardial infarction. *NEJM* 1977;297(20):1091-6.
21. Clagett GP, Reisch JS. Prevention of venous thromboembolism in general surgical patients. Results of meta-analysis. *Annals of Surgery* 1988;208(2):227-40.

22. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineburg HV, Mosteller F. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 1994;271(9):698-702.
23. Colditz GA, Halvorsen KT, Goldhaber SZ. Randomized clinical trials of transdermal nitroglycerin systems for the treatment of angina: A meta-analysis. *American Heart Journal* 1988;116(1Pt1):174-80.
24. Colditz GA, Tuden RL, Oster G. Rates of venous thrombosis after general surgery: Combined results of randomised clinical trials. *Lancet* 1986;2(8499):143-6.
25. Counsell C, Salinas R, Warlow C, Naylor R. Patch v no patch in carotoid surgery. *The Cochrane Library* 1996;Issue 3.
26. Counsell C, Sandercock P. LMWH/heparinoids in acute stroke. *The Cochrane Library* 1996;Issue 3.
27. Counsell C, Warlow C, Naylor R. Different patches in carotoid surgery. *The Cochrane Library* 1996;Issue 3.
28. Crowley P. Corticosteroids prior to preterm delivery. *The Cochrane Library* 1996;Issue 3.
29. Crowther C, Alfirevic Z. Antenatal TRH prior to preterm delivery. *The Cochrane Library* 1996;Issue 3.
30. Dardennes R, Even C, Bange F, Heim A. Comparison of Carbamazepine and Lithium in the Prophylaxis of Bipolar Disorders: A meta-analysis. *British Journal of Psychiatry* 166:378-381.
31. Daures JP, Schved JF, MoMAs I, Gril JC, Azoulay P, Gremy F. Meta analyse sur les essais randomises comparant les resultats des heparines de bas poids moleculaires a ceux des heparines fractionnees dans la prevention des thromboses veineuses profondes. *Rev. Epidem. et Sante* 7:363-369A.
32. Daya S. Comparison of FSH and HMG in IVF. *The Cochrane Library* 1996;Issue 3.
33. Daya S. Efficacy of progesterone support for pregnancy in women with recurrent miscarriage: A meta-analysis of controlled trials. *British Journal of Obstetrics and Gynaecology* 1989;96(3):275-80.

34. Daya S. Efficacy of progesterone support in the luteal phase following in-vitro fertilization and embryo transfer: meta-analysis of clinical trials. *Human Reproduction* 1988;3(6):731-4.
35. Dexter F, Tinker JH. Comparisons between desflurane and isoflurane or propofol on time to following commands and time to discharge: A meta- analysis. *Anesthesiology* 1995;83(1):77-82.
36. Duggan L, Brylewski J. Antipsychotics, learning diasb& Sz. *The Cochrane Library* 1996;Issue 3.
37. Duley L, Gulmezoglu AM, Henderson-Smart DJ. Anticonvulsants for pre-eclampsia. *The Cochrane Library* 1996;Issue 3.
38. Dush DM, Hirt ML, Schroeder HE. Self-statement modification in the treatment of child behaviour disorders: A meta-analysis. *Psychological Bulletin* 1989;106(1):97-106.
39. Edmonson JH., Su J, Krook JE. Treatment of Ovarian Cancer in Elderly Women. *Cancer Supplement* 1993;71(2):615-617.
40. Enkin MW, Wilkinson C. Absorbable staples for Caesarean section. *The Cochrane Library* 1996;Issue 3.
41. Fanning J, Bennett TZ, Hilgers RD. Meta-analysis of Cisplatin, Doxorubicin, and Cyclophosphamide Versus Cisplatin and Cyclophosphamide Chemotherapy of Ovarian Carcinoma. *Obstetrics & Gynecology* 1992;80:954-60.
42. Fraser WD, Krauss I, Brisson-Carrol G, Thornton J, breart G. Amniotomy to shorten spontaneous labour. *The Cochrane Library* 1996;Issue 3.
43. Fries SE, Levinton C, Naylor CD, Chen E, Christakis GT, Goldman BS. Optimal antithrombotic therapy following aortocoronary bypass: A Meta-analysis. *Eur J Cardio-thorac Surg* 1993;7:169-180.
44. Furberg CD, Bell RL. Effect of beta-blockers therapy on recurrent nonfatal myocardial infarction. *Circulation* 1983;67(6Pt2):I83-5.
45. Gateby PA, Cameron K, Simes RJ, Adelstein S, Bennett MJ, Jansen RPS, Shearman R, Stewart GJ, Whittle M, Doran TJ. Treatment of Recurrent

Spontaneous Abortion by Immunization With Paternal Lymphocytes: Results of a Controlled Trial. *American Journal of Reproductive Immunology* 1993;29:88-94.

46. Gent M, Roberts RS. A meta-analysis of the studies of dihydroergotamine plus heparin in the prophylaxis of deep vein thrombosis. *Chest* 1986;89(5):396S-400S.
47. Gillespie WJ, Henry DA, O'Connell D, Robertson J. Vitamin D in fracture prevention. *The Cochrane Library* 1996;Issue 3.
48. Glanz M, Klawansky S, Stason W, Berkey C, Shah N, Phan H, Chalmers TC. Biofeedback Therapy in Poststroke Rehabilitation: A Meta-analysis of the Randomized Controlled Trials. *Arch Phys Med Rehabil* 1995;76:508-515.
49. Glasziou PP, Mackerras DEM. Vitamin A supplementation in infectious diseases: A meta-analysis. *BMJ* 1993;306:366-370.
50. Golkstein P, Berrier J, Rosen S, Sacks HS, Chalmers TC. A meta-analysis of randomized control trials of progestational agents in pregnancy. *British Journal of Obstetrics and Gynecology* 1989;96(3):265-74.
51. Gore SM, Oldham FA. Randomized trials of high-versus-low-dose steroids in renal transplantation. Does the evidence favour a consensus? *Transplantation* 1986;41(3):319-27.
52. Gotzsche PC, Gjorup I, Bonnen H, Brahe NE, Becker U, Burcharth F. Somatostatin vs placebo in bleeding oesophageal varices: Randomised trial and meta-analysis. *BMJ* 1995;310(6993):1495-8.
53. Gotzsche PC. Patients preference in indomethacin trials: An overview. *Lancet* 1989;1(8629):88-91.
54. Grant A. Elective vs selective Caesarean section. *The Cochrane Library* 1996;Issue 3.
55. Grant A. The choice of suture materials and techniques for repair of perinatal trauma: An overview of the evidence from controlled trials. *British Journal of Obstetrics and Gynecology* 1989;96:1281-1289.
56. Graves P. Malaria vaccines. *The Cochrane Library* 1996;Issue 3.
57. Gulmezoglu AM, Hofmeyr GJ. Bedrest for impaired fetal growth. *The Cochrane Library* 1996;Issue 3.

58. Gulmezoglu AM, Hofmeyr GJ. Calcium channel blockers for impaired fetal growth. The Cochrane Library 1996;Issue 3.
59. Haines SJ, Walters BC. Antibiotic Prophylaxis for Cerebrospinal Fluid Shunts: A Meta-analysis. Neurosurgery 1994;34(1):87-92.
60. Hauth JC, Goldenberg RL, Parker CR, Cutter GR, Cliver SP. Low-dose aspirin: Lack of association with an increase in abruptio placentae or perinatal mortality. Obstetrics & Gynecology 1995;85(6):1055-1058.
61. Haynes RB, McKibbin KA, Kanani R. Adherence with medications. The Cochrane Library 1996;Issue 3.
62. Hedrick SC, Koepsell TD, Inui T. Meta-analysis of home-care effects on mortality and nursing-home placement. Medical Care 1989;27(11):1015-26.
63. Held PH, Yusuf S, Furberg CD. Calcium channel blockers in acute myocardial infarction and unstable angina: An overview. BMJ 1989;299(6709):1187-9.
64. Henderson WG, Goldman S, Copeland J, Moritz TE, Harker LA. Antiplatelet or anticoagulant therapy after coronary artery bypass surgery: A meta-analysis of clinical trials. Ann Intern Med 1989;111(9):743-50.
65. Henderson-Smart DJ, Steer P. Methylxanthine treatment for apnea of prematurity. The Cochrane Library 1996;Issue 3.
66. Henry DA, O'Connell DL. Effects of fibrinolytic inhibitors on mortality from upper gastrointestinal haemorrhage. BMJ 1989;298(6681):1142-6.
67. Heyland DK, Cook DJ, Jaeschke R, Griffith L, Lee HN, Guyatt GH. Selective Decontamination of the Digestive Tract: An overview. Chest 1994;105(4):1221-29.
68. Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction. Arch Intern Med 1989;149:2694-2698
69. Hodnett ED. Alternative versus conventional delivery settings. The Cochrane Library 1996;Issue 3.

70. Hodnett ED. Support for socially disadvantaged mothers. The Cochrane Library 1996;Issue 3.
71. Hofmeyr GJ. Abdominal decompression. The Cochrane Library 1996;Issue 3.
72. Hofmeyr GJ. Amnioinfusion for meconium liquor in labour. The Cochrane Library 1996;Issue 3.
73. Hofmeyr GJ. Cephalic version by postural management. The Cochrane Library 1996;Issue 3.
74. Hofmeyr GJ. External cephalic version at term. The Cochrane Library 1996;Issue 3.
75. Hofmeyr GJ. Maternal oxygen for fetal distress. The Cochrane Library 1996;Issue 3.
76. Hofmeyr GJ. Planned Caesarean section for breech. The Cochrane Library 1996;Issue 3.
77. Hofmeyr GJ. Prophylactic versus amniofusion. The Cochrane Library 1996;Issue 3.
78. Holme I. Drug treatment of mild hypertension to reduce the risk of CHD: Is it worth-while? *Statistics in Medicine* 1988;7(11):1109-20.
79. Hopfenmuller W. Nackweis der therapeutischen Wirksamkeit eines Ginkgo biloba-Spezialextrakes: Meta-Analyse von 11 klinischen Studien bei Patienten mit Hirnleistungsstörungen im Alter. *Arzneim-Forsch* 1994;44(9):1005-13.
80. Hughes E, Collins J, Vandekerckhove, P. Bromocriptine, unexplained infertility. The Cochrane Library 1996;Issue 3.
81. Hughes E, Collins J, Vandekerckhove P. GnRHa/hMG/FSH for PCOS. The Cochrane Library 1996;Issue 3.
82. Hughes E, Fedorkow DM, Daya S, Sagle MA, van de Kopple P, Collins JA. The routine use of gonadotropin-releasing hormone agonists prior to in vitro fertilization and gamete intrafollicular transfer: A meta-analysis of randomized controlled trials. *Fertility and Sterility* 1992;58(5):888-96.

83. Infante-Rivard C, Esnaola S, Villeneuve JP. Role of endoscopic variceal sclerotherapy in the long-term management of variceal bleeding: A meta-analysis. *Gastroenterology* 1989; 96(4):1087-92
84. Insua JT, Sacks HS, Lau T, Lau J, Reitman D, Pagano D, Chalmers TC. Drug Treatment of Hypertension in the Elderly: A Meta-analysis. *Ann Intern Med* 1994;121:355-362 .
85. Janicak PG, Davis JM, Gibbons RD, Ericksen S, Chang S, Gallagher P. Efficacy of ECT: a meta-analysis *American Journal of Psychiatry* 1985;142(3):297-302.
86. Kauffman NA, Herman CP, Polivy J. Hunger-induced Finickiness in Humans. *Appetite* 1995;24:203-218.
87. Kaufmann PG, Jacob RG, Ewart CK, Chesney MA, Muenz LR, Doub N, et al. Hypertension Intervention Pooling Project. *Health Psychology* 1988;7:209-24.
88. King JF, Grant A, Keirse MJ, Chalmers I. Beta-mimetics in preterm labour: An overview of the randomized controlled trials. *British Journal of Obstetrics and Gynecology* 1988;95(3):211-22.
89. Koch M, Capurso L, Dezi A, Ferrario F, Scarpignato C. Prevention of NSAID-Induced Gastroduodenal Mucosal Injury: Meta-analysis of Clinical Trials with Misoprostol and H2-receptor Antagonists. *Dig Dis* 1995;3(suppl1):62-74.
90. Koudstaal P. Secondary stroke prevention in AF: Part III. *The Cochrane Library* 1996;Issue 3.
91. Kramer MS. Isocaloric balanced protein supplementation. *The Cochrane Library* 1996;Issue 3.
92. Kramer MS. Maternal antigen avoidance as lactation. *The Cochrane Library* 1996;Issue 3.
93. Kramer MS. Nutritional advice in pregnancy. *The Cochrane Library* 1996;Issue 3.
94. Kramer MS. Maternal antigen avoidance as treatment. *The Cochrane Library* 1996; Issue 3.

95. Lensing AWA, Prins MH, Davidson BL, Hirsh J. Treatment of Deep Venous Thrombosis With Low-Molecular-Weight Heparins. *Arch Intern Med* 1995;155:601-607.
96. Lievre M, Leizorovicz A, Boisseol JP. Criteres intermediaires et criteres desubstitution dans le developpement des antiarythmiques. *Archives des Maladies du Coeur et des Vaisseaux* 1991;84(2):27-33.
97. Lycka BA. Postherpetic neuralgia and systemic corticosteroid therapy. Efficacy and safety. *International Journal of Dermatology* 1990;29(7):523-7.
98. MacRae HM, McLeod RS. Comparison of hemorrhoidal treatment modalities: A meta-analysis. *Diseases of the Colon & Rectum* 1995;38(7):687-94.
99. Mahomed K. Blood transfusion in sickle cell anaemia. *The Cochrane Library* 1996;Issue 3.
100. Mahomed K, Gulmezoglu AM. Maternal iodine supplements in areas of deficiency. *The Cochrane Library* 1996;Issue 3.
101. Mahomed K, Gulmezoglu AM. Vitamin D supplementation in pregnancy. *The Cochrane Library* 1996;Issue 3.
102. Mari JJ, Streiner D. Family intervention & Schizophrenia. *The Cochrane Library* 1996;Issue 3.
103. McGrath JJ, Soares KVS. Tardive dyskinesia and benzodiazepines. *The Cochrane Library* 1996;Issue 3.
104. Meunier F, Paesman M, Autier P. Value of Antifungal Prophylaxis with Antifungal Drugs Against Oropharyngeal Candidiasis in Cancer Patients. *Oral Oncol* 1994;30B(3):196-199.
105. Midgette AS, Wong JB, Beshansky JR, Porath A, Fleming C, Pauker SG. Cost-Effectiveness of Streptokinase for Acute Myocardial Infarction: A Combined Meta-analysis and decision analysis of the effects of infarct Location and of likelihood of infarction. *Med Decis Making* 1994;14:108- 117.
106. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: A quantitative review of primary prevention trials. *BM.* 1990;301(6747):309-14.

107. Mulrow CD, Mulrow JP, Linn WD, Aguilar C, Ramirez C. Relative efficacy of vasodilator therapy in chronic congestive heart failure. Implications of randomized trials. *JAMA* 1988;259(23):3422-6.
108. Naveau S, Perrier C, Mory B, Poynard T, Chaput JC. Hemostase endoscopique des hemorrhagies ulcereuses gastro-duodenales. *Gastroenterologie Clinique et Biologique* 1991;15(8-9):580-7.
109. Naylor CD, Detsky AS, O'Rourke K, Fonberg E. Does treatment with essential amino acids and hypertonic glucose improve survival in acute renal failure? A Meta-analysis. *Renal Failure* 1987;10(3-4):141-52.
110. Neilson JP. Routine early pregnancy ultrasound. *The Cochrane Library* 1996;Issue 3.
111. O'Brien BJ, Anderson DR, Goeree R. Cost-effectiveness of enoxaparin versus warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *Can Med Assoc J* 1994;150(7):1083-1172.
112. O'Connor GT, Buring JE, Yusuf S, Goldhaber SZ, Olmstead EM, Paffengarger RS, Hennekens CH. An overview of randomized trials of rehabilitation with exercise after myocardial infarction. *Circulation* 1989;80(2):234-44.
113. Ohlsson A. Treatments of preterm premature rupture of the membranes: a meta-analysis. *American Journal of Obstetrics & Gynecology* 1989;160(4):890-906.
114. Oldridge NB, Guyatt GH, Fischer ME, Rimm AA. Cardiac rehabilitation after myocardial infarction. Combined experience of randomized controlled trials. *JAMA* 1988;260(7):945-50.
115. Olliaro P, Mussano P. Amodiaquine for treatment of malaria. *The Cochrane Library* 1996;Issue 3.
116. Ottenbacher KJ, Cooper HM. Drug treatment of hyperactivity in children. *Development Medicine & Child Neurology* 1983;25(3):358-66.
117. Patel M, Gutzwiller F, Paccaud F, Marazzi A. A meta-analysis of acupuncture for chronic pain. *International Journal of Epidemiology* 1989;18(4):900-6.
118. Patten SB. Propranolol and Depression: Evidence from the Antihypertensive Trials. *Can J Psychiatry* 1990;35:257-259.

119. Perez-Escamilla R, Pollitt E, Lonnerdal B, Dewey KG. Infant Feeding Policies in Maternity Wards and Their Effect on Breast-Feeding Success: An Analytical Overview. *American Journal of Public Health* 1994;84:89-97.
120. Pignon JP, Poynard T, Naveau S, Sebag G, Chaput JC. Traitement du reflux gastro-oesophagien: analyse des essais randomisés en double aveugle. *Gastroenterologie Clinique et Biologique* 1987;11(10):668-80.
121. Prendiville W, Elbourne D, Chalmers I. The effects of routine oxytocic administration in the management of the third stage of labour: An overview of the evidence from controlled trials. *British Journal of Obstetrics and Gynaecology* 1988;95(1):3-16.
122. Radack K, Deck C. The effects of omega-3 polyunsaturated fatty acids on blood pressure: A methodologic analysis of the evidence. *Journal of the American College of Nutrition* 1989;8(5):376-85.
123. Radack K, Wyderski RJ. Conservative management of intermittent claudication. *Ann Intern Med* 1990;113(2):135-46.
124. Ramirez-Lassepas M, Cipolle RJ. Medical treatment of transient ischemic attacks: Does it influence mortality? *Stroke* 1988;19(3):397-400.
125. Renfrew MJ, Lang S. Breastfeeding and discharge times. *The Cochrane Library* 1996;Issue 3.
126. Renfrew MJ, Lang S. Feeding technique and breastfeeding. *The Cochrane Library* 1996;Issue 3.
127. Renfrew MJ, Lang S. Breastfeeding and early contact. *The Cochrane Library* 1996;Issue 3.
128. Richter JM, Colditz GA, Huse DM, Delea TE, Oster G. Cimetidine and adverse reactions: a meta-analysis of randomized clinical trials of short-term therapy. *American Journal of Medicine* 1989;87(3):278-84.
129. Schmader KE, Studenski S. Are current therapies useful for the prevention of postherpetic neuralgia? A critical analysis of the literature. *Journal of General Internal Medicine* 1989;4(2):83-9.
130. Schneider LS. Ginkgo-biloba-Extrakt bei peripheren arteriellen Verschlusskrankheiten: Meta-analysis von kontrollierten klinischen Studien. *Drugs* 1992;42(1):428-436.

- 131.Schneider LS, Olin JT. Overview of Clinical Trials of Hydergine in Dementia. Arch Neurol 1994;51:787-798.
- 132.Silagy C, Mant D, Fowler G, Lancaster T. Nicotine replacement therapy. The Cochrane Library 1996;Issue 3.
- 133.Souza KVS, McGrath JJ, Deeks JJ. Tardive dyskinesia and GABA agonist drugs. The Cochrane Library 1996;Issue 3.
- 134.Solimann S, Daya S, Collins J, Hughes EG. The role of luteal phase support in infertility treatment: A meta-analysis of randomized trials. Fertility and Sterility 1994;61(6):1068-1076.
- 135.Souza FGM, Goodwin GM. Lithium Treatment and Prophylaxis in Unipolar Depression: A Meta-analysis. British Journal of Psychiatry 1991;158:666-675.
- 136.Stroke Unit Trialists' Collaboration. Stroke Units. The Cochrane Library 1996;Issue 3.
- 137.Teo KK, Yusuf S. Role of Magnesium in Reducing Mortality in Acute Myocardial Infarction: A Review of the Evidence. Drugs 1993;46(3):347-359.
- 138.Thacker SB. Quality of controlled clinical trials. The case of imaging ultrasound in obstetrics: a reievw. British Journal of Obstetrics and Gynacology 1985;92(5):437-44.
- 139.Thomas JA, McIntosh JM. Are incentive Spirometry, Intermittent Positive Pressure Breathing, and Deep Breathing Exercises Effective in the Prevention of Postoperative Pulmonary complications, After Upper Abdominal Surgery? A Systematic Overview. Physical Therapy 1994;74(1):3-10.
- 140.Tine F, Liberati A, Craxi A, AlMASio P, Pagliaro L. Interferon treatment in patients with chronic hepatitis B: a meta-analysis of the published literature. Journal of Hepatology 1993;18(2):154-160.
- 141.Trivedi MH, Rush J. Does Placebo Run-In or a Placebo Treatment Cell Affect the EfficacyofAntidepressantMedications?Neuropsychopharmacology 1994;11(1):33-43.
- 142.Vanderkerckhove P, Lilford R, Hughes E. Anti-oestrogens for oligospermia. The Cochrane Library 1996;Issue 3.

- 143.Vanderkerckhove P, Lilford R. Androgens for oligospermia. The Cochrane Library 1996;Issue 3.
- 144.Velanovich V. Crystalloid versus colloid fluid resuscitation: a meta-analysis of mortality. Surgery 1989;105(1):65-71.
- 145.Vito Agosti MSW. The Efficacy of Treatments in Reducing Alcohol Consumption: A Meta-analysis. The International Journal of the Addictions 1995;30(8):1067-1077.
- 146.Warshafsky S, Kamer RS, Sivak SL. Effect of garlic on total serum cholesterol. A meta analysis. Ann Intern Med 1993;119(7Pt1):599-60.
- 147.Whitelaw A. Treatment of IVH by CSF tapping. The Cochrane Library 1996;Issue 3.
- 148.Williams RL Chalmers TC, Stange KC, Chalmers FT, Bowlin SJ. Use of Antibiotics in Preventing Recurrent Acute Otitis Media and in Treating Otitis Media with Effusion: A Meta-analytic Attempt to Resolve the Brouhaha. JAMA 1993;270(11):1344-135.
- 149.Wilson APR., Shrimpton S, Jaderberg M. A meta-analysis of the use of amoxycillin-clavulanic acid in surgical prophylaxis. Journal of Hospital Infection 1992;22(supplA):9-21.
- 150.Young GL, Jewell MD. Creams to prevent striae gravidarum. The Cochrane Library 1996;Issue 3.
- 151.Yusuf S. An overview of the clinical trials of agents (other than beta-blockers) that potentially limit myocardial infarct size. Journal of Cardiovascular Pharmacology. 1988;12(Suppl1):S48-5.