IMPACT OF GREY LITERATURE ON SYSTEMATIC REVIEWS OF RANDOMIZED TRIALS

Sally Hopewell

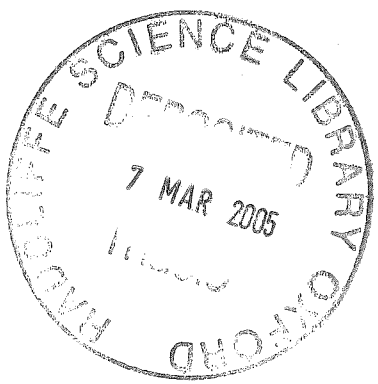
Wolfson College

University of Oxford

Doctor of Philosophy Thesis

Trinity Term

2004



ABSTRACT

Impact of grey literature on systematic reviews of randomized trials

Sally Hopewell, Wolfson College, D.Phil. thesis, Trinity Term 2004

Systematic reviews are increasingly important in evidence-based medicine. Their validity is dependent on the underlying data, thus any decisions about the eligibility of trials based on where they are published, for example excluding trials from the grey literature, could bias their results and conclusions. However, the identification and inclusion of trials published in the grey literature can be time consuming and difficult. This thesis assesses the actual and potential impact of including trials reported in the grey literature in systematic reviews of healthcare interventions. A review of existing knowledge in this area was first carried out. Four substantial projects were then conducted.

The first of these projects assessed the actual impact of grey literature in meta-analyses published in the *Cochrane Database of Systematic Reviews*. This showed that published trials may produce larger treatment effects than trials identified in the grey literature (grey trials). There was no difference in methodological quality between published and grey trials; this was often unclear or not reported in both sources. The second project assessed the potential impact of grey literature in systematic reviews. Practical issues in locating and including grey trials were identified, namely the quantity and quality of information available. The third project built on these findings, assessing the quality of trials reported in conference proceedings and whether this improved over time. Important limitations were identified including a lack of numerical data and the preliminary nature of the results presented; these appear to be worse in recent years. The final project assessed how trial information reported in a conference abstract differs from its full publication. Major changes between abstract and full publication were identified, in particular details of trial results and participants included and analysed in the trial. It reinforces the preliminary nature of some trials reported in conference proceedings.

This thesis confirms that the exclusion of trials reported only in the grey literature remains a threat to the validity and reliability of systematic reviews. Systematic reviewers must, therefore, search for and include trials from the published and grey literature. Failure to do so may result in the results and conclusions of that review being biased and unreliable.

ACKNOWLEDGEMENTS

Firstly, I should like to express my thanks to my supervisor Dr Mike Clarke, who has been extremely supportive, offering constant advice and guidance throughout this thesis.

I should like to acknowledge the contributions made by Matthias Egger, Steve McDonald and Mike Clarke for their assistance in conducting the Cochrane Methodology review presented as part of this thesis. I should like to thank Sue Mallett for providing additional data for this thesis, Phil Alderson, Mike Clarke, Anne Eisinga, Liz MacKinnon, Steve McDonald, Philippa Middleton, Roberta Scherer, Nancy Owen and Lisa Askie for their assistance with data extraction, Jon Deeks for statistical advice and Anne Eisinga for proof reading. I should also like to thank colleagues at the UK Cochrane Centre for their support and encouragement and also thank the UK Cochrane Centre for funding my time at the University of Oxford.

Finally, I should like to thank my friends and family for putting up with me and keeping me sane through the past three years!

TABLE OF CONTENTS

CHAPTER 2 : BACKGROUND AND RATIONALE 1

Introduction 1

What is a systematic review? 1

A brief history of systematic reviews 2

Rationale for systematic reviews 4

Principles of a systematic review 6

Bias in systematic reviews and meta-analyses 8

Publication bias 8

Time- lag bias 11

Language bias 13

Citation bias 13

Duplication bias 14

Bias and methodological quality 14

Minimising bias through a comprehensive search 16

Differences between health and social sciences 20

What is grey literature? 20

Sources of grey literature 22

Use of grey literature in systematic reviews 24

Impact of grey literature in social sciences 28

Impact of grey literature in health sciences 30

Problems in searching for grey literature 31

Difficulties of including grey literature 33

RESEARCH OUTLINE 36

Aims 36

Objectives 36

Research approach 37

CHAPTER 3 : SYSTEMATIC REVIEW OF THE LITERATURE 40

Introduction 40

Objectives 41

Studies and data to be considered 41
Types of study 41
Data to be extracted from these studies 41
Main outcome measure 41
Search strategy for identification of studies 42
Methods of the review 43
Results 44
Description of studies 44
Characteristics of the meta-analyses 45
Characteristics of the trials included in the meta-analyses 48
Language of publication of the trials 50
Type of grey literature 50
Methodological quality of the randomized trials included in the meta-analyses ..51
Outcome measures for grey and published trials 52
Discussion 54
Limitations of the existing evidence 58
CHAPTER 4 : ACTUAL IMPACT OF INCLUDING GREY LITERATURE IN META-ANALYSES OF RANDOMIZED TRIALS 60
Introduction 60
Objectives 60
Methods 61
Source of data 61
Inclusion criteria 61
Data extracted from each systematic review 61
Data extracted from each meta-analysis 63
Data extracted from each trial 64
Data analysis 66
Results 68
Description of all 130 systematic reviews 68
Description of all 25 systematic reviews included in the main analysis 72
Trials included in the meta-analyses 76

Statistical model and type of effect estimate used in the meta-analyses 77
Participants included in the meta-analyses 77
Sources of the trials included in the meta-analyses 78
Methodological quality of the trials included in the meta-analyses 79
Estimates of treatment effects for published and grey trials 82
Estimates of treatment effects for methodological quality 89
Discussion 92
Limitations of the existing evidence 100
CHAPTER 5 : POTENTIAL IMPACT OF INCLUDING GREY LITERATURE IN SYSTEMATIC REVIEWS 101
Introduction 101
Objectives 102
Source of data 102
Methods 103
Handsearching 103
Assessing full publication 104
Assessment of trial characteristics 105
Searching for possible systematic reviews 105
Results of the pilot study 106
Full publication 107
Trial characteristics 108
Identifying possible systematic reviews 110
Results of the main study 111
Full publication 112
Trial characteristics 114
Identifying possible systematic reviews 114
Discussion 116
Limitations of the existing evidence 119
CHAPTER 6 : QUALITY OF TRIALS REPORTED IN THE GREY LITERATURE 120
Introduction 120

Objectives 121
Source of data 121
Methods 122
Handsearching 122
Assessing the quality of the abstract 122
Data extraction and analysis 126
Results 127
Part 1: Quality of published and unpublished trials reported in abstracts 127
Reporting of authors 127
Reporting of objectives 127
Reporting of study design 128
Reporting of study quality 129
Reporting of participants 130
Reporting of interventions 130
Reporting of outcomes 131
Reporting of results 133
Reporting of conclusions 136
Part 2: Quality of trials reported in abstracts over time 136
Reporting of authors 138
Reporting of objectives 138
Reporting of study design 139
Reporting of study quality 139
Reporting of participants 140
Reporting of interventions 140
Reporting of outcomes 141
Reporting of results 143
Reporting of conclusions 145
Discussion 146
Limitations of the existing evidence 150
CHAPTER 7 : COMPARING TRIALS REPORTED IN THE GREY LITERATURE WITH THEIR FULL PUBLICATION 151

Introduction 151
Objectives 151
Methods 152
Source of data 153
Data extraction 153
Data analysis 154
Results 157
Part 1: Comparing trials reported as conference abstracts with their subsequent full publication 157
Reporting of general characteristics 157
Reporting of authors 161
Reporting of objectives 162
Reporting of study design 163
Reporting of study quality 164
Reporting of participants 165
Reporting of interventions 166
Reporting of outcomes 166
Reporting of results 171
Reporting of conclusions 173
Part 2: Comparing trials reported as conference abstracts with the abstract of their subsequent full publication 174
Reporting of objectives 175
Reporting of study quality 176
Reporting of participants 176
Reporting of interventions 176
Reporting of outcomes 177
Reporting of results 180
Reporting of conclusions 181
Discussion 184
CHAPTER 8 : DISCUSSION, IMPLICATIONS AND CONCLUSIONS 192
Introduction 192

Discussion of main findings....192
Impact of randomized trials reported in the grey literature....193
Quality of randomized trials reported in the grey literature....201
Reporting of randomized trials in the grey literature....204
Implications for systematic reviewers....210
Implications for trialists....211
Implications for users of trial evidence....212
Implications for future research....213
Conclusions....215
REFERENCES....218
APPENDIX 1: Publications relating to this thesis....232
APPENDIX 2: Analysis of estimates of treatment effect....233
APPENDIX 3: Data extraction form....234
APPENDIX 4: Abstract comparing trials reported as abstracts and full papers....235

TABLE OF TABLES

Table 2:1: Sources of bias 8

Table 2:2: Citations for grey literature in 513 Cochrane reviews 25

Table 3:1: Characteristics of the included studies 46

Table 3:2: Characteristics of the meta-analyses 48

Table 3:3: Characteristics of the trials included in the meta-analyses 49

Table 3:4: Methodological quality of the trials included in the meta-analyses 52

Table 3:5: Summary outcome measures for grey and published trials 53

Table 4:1: Country of contact author of the review 68

Table 4:2: Details of included studies (n=130) 70

Table 4:3: Details of excluded studies and studies awaiting assessment mentioned in the reviews (n=130) 71

Table 4:4: Review titles and the number of trials per meta-analysis 74

Table 4:5: Details of included studies (n=25) 75

Table 4:6: Details of excluded studies and studies awaiting assessment mentioned in the reviews (n=25) 76

Table 4:7: Trials included in the meta-analyses 77

Table 4:8: Statistical model and type of effect estimate used in the meta-analyses 77

Table 4:9: Participants included in the meta-analyses 78

Table 4:10: Allocation concealment score 80

Table 4:11: Methodological quality of the trials included in the meta-analyses 82

Table 4:12: Odds ratios for published trials versus overall trials 83

Table 5:1: Number of trials reported as abstracts in the Australian and New Zealand Journal of Medicine and full publication of their results 108

Table 5:2: Characteristics of the trials reported in the Thoracic Society of Australia and New Zealand abstracts 109

Table 5:3: Trials reported in the Thoracic Society of Australia and New Zealand abstracts for which there was a possible Cochrane review 110

Table 5:4: Number of trials reported as abstracts at the American Society of Clinical Oncology conference and full publication of their results 111

Table 5:5: Characteristics of the trials reported as abstracts at the American Society of Clinical Oncology conference 114

Table 5:6: Trials reported as abstracts at the American Society of Clinical Oncology conference for which there was a possible Cochrane review 115

Table 6:1: Reporting of author correspondence details, objectives and study design. 128

Table 6:2: Reporting of study quality 129

Table 6:3: Reporting of participants and interventions 130

Table 6:4: Reporting of number of participants randomized and analysed 132

Table 6:5: Reporting of trial status and intention- to-treat analysis 133

Table 6:6: Reporting of results 134

Table 6:7: Reporting of conclusions 136

Table 6:8: Number of trials reported as abstracts and site of cancer 137

Table 6:9: Reporting of author correspondence details, objectives and study design. 138

Table 6:10: Reporting of study quality 139

Table 6:11: Reporting of participants and interventions 140

Table 6:12: Reporting of number of participants randomized and analysed 142

Table 6:13: Reporting of trial status and intention to treat analysis 143

Table 6:14: Reporting of results 144

Table 6:15: Reporting of conclusions 145

Table 7:1: Reporting of trial name or identifier 162

Table 7:2: Reporting of date of trial 163

Table 7:3: Reporting of "trial" in title 163

Table 7:4: Reporting of "randomized" in title 164

Table 7:5: Reporting of method of allocation concealment 164

Table 7:6: Reporting of method of blinding 165

Table 7:7: Reporting of intervention groups 166

Table 7:8: Reporting of trial status 167

Table 7:9: Reporting of number of participants randomized and analysed 168

Table 7:10: Changes in the number of participants reported 170

Table 7:11: Reporting of intention-to-treat analysis 170

Table 7:12: Reporting of adverse effects 171

Table 7:13: Type of result reported 173

Table 7:14: Reporting of trial name or identifier 175

Table 7:15: Reporting of date of trial 176

Table 7:16: Reporting of intervention groups 177

Table 7:17: Reporting of trial status 178

Table 7:18: Reporting of intention-to-treat analysis 180

Table 7:19: Reporting of adverse effects 180

Table 7:20: Summary of level of agreement between conference abstracts, full publications and abstracts of the full publications.....183

TABLE OF FIGURES

Figure 1:1: A meta-analysis is an optimal part of a systematic review 7

Figure 1:2: Source of primary reference for trials cited by grey literature alone included in Cochrane reviews in 2001 27

Figure 2:1: Type of grey literature included in the meta-analyses 51

Figure 3:1: Collaborative Review Group and number of systematic reviews 69

Figure 3:2: Process for identifying eligible systematic reviews for inclusion in the main analysis 73

Figure 3:3: Source of trials included in the meta-analyses 79

Figure 3:4: Ratio of estimates of treatment effect from grey trials compared to published trials 85

Figure 3:5: Ratio of estimates of treatment effect from published trials supplemented with personal communication compared to published trials 87

Figure 3:6: Ratio of estimates of treatment effect from grey trials compared to published trials supplemented with personal communication 88

Figure 3:7: Ratio of estimates of treatment effect from trials with adequate versus unclear concealment of allocation sequence 90

Figure 3:8: Ratio of estimates of treatment effect from trials with adequate versus unclear/inadequate generation of allocation sequence 91

Figure 3:9: Ratio of estimates of treatment effect from double blind trials versus those from other trials 92

Figure 3:10: Meta-analysis of studies assessing the methodological quality of randomized trials 98

Figure 4:1: Time to publication 113

Figure 5:1: Checklist for reporting a randomized trial as an abstract..... 125

Figure 6:1: Flow diagram showing the study eligibility criteria 156

Figure 6:2: Time to full publication..... 157

Figure 6:3: Number of new authors listed in the full publication..... 159

Figure 6:4: Source of funding..... 161

CHAPTER 1: BACKGROUND AND RATIONALE

Introduction

The aim of this thesis is to examine the actual and potential impact of including trials found in the grey literature in systematic reviews. This opening chapter identifies what is grey literature and outlines existing evidence on why it is important to include trials reported in the grey literature in systematic reviews in order to help minimise the effects of bias. The end of this first chapter provides a detailed description of the research objectives and the approach taken. Chapters two to six will then present the thesis results and assess their validity. The concluding chapter will discuss the implications of these findings and suggest ways in which to address these issues along with possibilities for future research.

What is a systematic review?

Healthcare providers, researchers and policy makers are inundated with vast and unmanageable amounts of information (Mulrow, 1994). This volume of information is constantly changing and expanding and, in many areas, it would be impossible for individuals to read, critically evaluate and synthesise the current state of knowledge, let alone keep up to date (Egger et al, 2001a).

Reviews of this information are a valuable way of digesting and keeping up to date with the current state of knowledge. Narrative reviews seek to reduce the large quantities of information into more manageable and usable forms. Unfortunately,

however, narrative reviews are often subjective and can be of poor quality. Strategies for identifying and selecting information are also rarely defined. A study by Mulrow (1987) assessed the quality of 50 reviews published in four major medical journals between June 1985 and June 1986. Her study found that medical reviews, at that time, did not routinely use scientific methods to identify, assess and synthesise information. This study was repeated ten years later to determine if the quality of review articles had improved. It was found that although there have been improvements, there was still wide variation in the methodological quality of review articles (McAlister et al, 1999).

A systematic review differs from a narrative review of the literature in that there is a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research and to collect and analyse data from studies that are included in the review (Alderson et al, 2003).

A brief history of systematic reviews

James Lind's book published in 1753 called the *Treatise of Scurvy* is thought to contain one of the earliest examples of a systematic review. Lind's book contained a "... critical and chronological view..." of what has been published about scurvy, which was killing thousands of people every year, as well as one of the earliest accounts of a fair comparison of different medical treatments. Lind showed that oranges and lemons were more effective than other treatments for scurvy (Lind, 1753).

In 1904, Karl Pearson published a report in the *British Medical Journal* that combined data on enteric fever inoculations from data provided by the British Army in India and South Africa. He is thought to be the first medical researcher to combine data from a number of different studies using a formal statistical technique (Pearson, 1904).

Much of the early work in developing the formal methods of carrying out a systematic review was in social sciences research, in particular education and psychology. Glass and colleagues (1981) were instrumental in the development of formal methods for systematic reviews in the area of social sciences research. Glass was also the first person to use the term "meta-analysis" to describe the statistical technique for combining data from individual studies (Glass, 1976).

In an influential book published in 1972, Archie Cochrane first drew attention to the fact that people who want to make more informed decisions about health care do not have access to reliable reviews of the available evidence (Cochrane, 1972). Tom Chalmers was also influential in recognising that information from randomized trials needed to be systematically and cumulatively synthesised and is considered to be one of the founders and developers of systematic review methodology in health care.

The Cochrane Collaboration was developed in response to Archie Cochrane's call for systematic, up-to-date reviews of all relevant randomized trials of health care. The Cochrane Collaboration grew from a pilot study in 1974 in the area of perinatal medicine, where randomized trials were systematically identified and assembled in a

trials register. By 1985 the register contained more than 3,500 reports of controlled trials and led to the preparation of around 600 systematic reviews in pregnancy and childbirth by the late 1980s (Antes and Oxman, 2001).

A well-reported example of the importance of systematically reviewing data from randomized trials in health care is the administration of corticosteriods before premature labour. Twelve randomized trials, involving over 3,000 participants, were conducted between 1972 and 1980. However, it was not until a systematic review of these trials was carried out in 1989 that obstetricians realised the treatment was so effective in reducing the risk of respiratory distress syndrome in neonates and, as a result, tens of thousands of premature babies had suffered complications or died unnecessarily (Crowley et al, 1990). The work carried out in the area of perinatal medicine became the precursor of The Cochrane Collaboration which is now an international organisation carrying out systematic reviews of the effects of healthcare interventions across the world.

Rationale for systematic reviews

The rationale for systematic reviews is firmly grounded in the need to refine large amounts of unmanageable information into a usable format (Mulrow, 1994). Systematic reviews are an efficient scientific technique and, although arduous and time consuming, a review is usually quicker and less costly than embarking on a new study. In some cases, systematic reviews may, in fact, demonstrate the lack of adequate evidence and thus identify where further research is needed.

The results of a systematic review may be generalisable across different settings and might show consistency of the treatment effect across different interventions (for example varying drug dosages). Combining information from a number of studies of the same intervention can also help explain data inconstancy and conflicts in data between studies. They also have increased power and increased precision compared to a single study. Ultimately, systematic reviews will allow healthcare providers, policy makers and others to integrate the best available evidence into practice.

Cochrane systematic reviews are increasingly being cited in national and international guidelines and policies and provide an encouraging example of how systematic reviews of the best available evidence are being used to inform decisions about policy, practice and new research. This is demonstrated in a recent study, which found that Cochrane reviews are widely cited in a variety of reports such as those published by the National Health Service (NHS) Health Technology Assessment Programme in the UK, the NHS Centre for Reviews and Dissemination's bulletin *Effective Health Care*, the Finnish *Evidence-Based Medicine Guidelines* for general practitioners, and the BMJ Publishing Group's publication *Clinical Evidence* (Holmes et al, 2001).

The Reproductive Health Library (RHL) produced by the World Health Organisation is another example of where systematic reviews have been used to inform and change practice. The RHL provides up-to-date systematic reviews focusing on evidence-based solutions to reproductive health problems in developing countries and is built around a

core set of Cochrane systematic reviews in the field of reproductive health (www.update-software.com/rhl/).

A further example is the National Institute for Clinical Excellence (NICE), which was set up in April 1999 (www.nice.org.uk). NICE is part of the National Health Service in the UK and its role is to provide patients, health professionals and the public with authoritative, robust and reliable guidance on current "best practice" by conducting systematic reviews of the available evidence. The guidance covers both individual health technologies (including medicines, medical devices, diagnostic techniques and procedures) and the clinical management of specific conditions.

Principles of a systematic review

Systematic reviews should have a well-formulated review question with predefined study inclusion and exclusion criteria. The methods on how to locate and select relevant studies for inclusion in the review should also be explicit and predefined, as should the methods of assessing study quality, extracting the data and analysing and presenting the results. One of the fundamental ways to minimise introducing bias into a systematic review is to state explicitly the hypotheses and the methodological approaches that will be used in conducting the review before undertaking it, and without prior knowledge of the data (Silagy et al, 2002). In the case of Cochrane systematic reviews, all protocols are peer reviewed and published electronically in *The Cochrane Library*. This also allows the authors of the review to obtain additional feedback before proceeding to data collection and analysis.

A meta-analysis is the statistical procedure that integrates the results of several independent studies, which are considered to be combinable. A well-conducted meta-analysis allows a more objective appraisal of the evidence and provides a more precise estimate of the treatment effect. It may also explain heterogeneity (variation) between the results of individual studies (Egger et al, 1997). A meta-analysis should be seen as an optional part of a systematic review (Figure 1.1). It might not be appropriate to conduct a meta-analysis, when conducting a systematic review, if there are too few studies estimating the treatment effect or if there are differences in the study characteristics (for example, clinical, statistical or methodological heterogeneity between studies) that are likely to substantially affect the outcome. It may also not be possible to carry out a meta-analysis when the outcomes from individual studies have been measured in different ways or if data for a study are missing or incomplete.

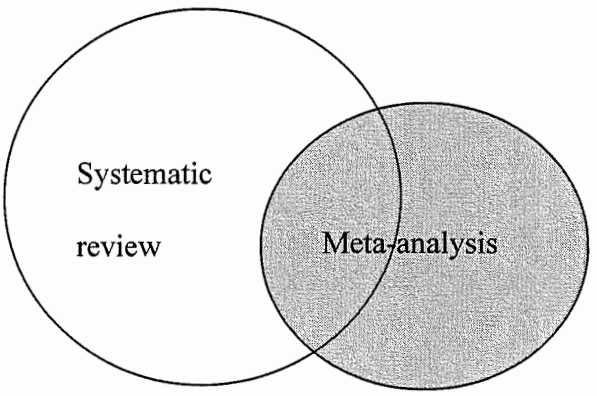


Figure 1:1: A meta-analysis is an optional part of a systematic review

In some cases a meta-analysis may be carried out outside a systematic review. However, the results of such a meta-analysis may be biased, as explicit systematic methods of identifying, selecting and critically appraising relevant studies may not have been deployed (Eysenck, 1994).

Bias in systematic reviews and meta-analyses

The validity of the results of a systematic review is highly dependent on the results of the studies that it includes. There are a number of different ways in which bias can be introduced when conducting a systematic review (Table 1.1).

Table 1:1: Sources of bias

Type of bias	Definition
Publication bias	Studies with statistically significant results are more likely to be published than those with non-statistically significant or null results.
Time-lag bias	Studies with statistically significant results are more likely to be stopped earlier than originally planned and published quicker.
Language bias	Studies with statistically significant results are more likely to be published in English.
Duplication bias	Studies with statistically significant results are more likely to be published more than once.
Citation bias	Studies with statistically significant results are more likely to be cited by others.

Publication bias

Bias is most commonly introduced into systematic reviews due to the reporting or dissemination of the trials included in it and their research findings. It has been known for some time that the publication or non-publication of research findings is dependent on the nature and direction of the results (Rosenthal, 1979). The term publication bias

has been widely used to describe where research with significant results are more likely to be submitted and published than research with null or non-significant results (Dickersin, 1997). Possibly the earliest examples of the existence of publication bias are the studies by Sterling et al (1959) and Smart (1964). Sterling et al and Smart identified that the results of the majority of studies published in four major psychology journals were statistically significant. They concluded that studies with non-significant results might be under-represented in the published literature.

A number of studies have examined the publication of studies of biomedical research from when they were initiated to publication of their results. In a study by Easterbrook and colleagues (1991), 285 clinical research studies initially approved by the Central Oxford Research Ethics Committee between 1984 and 1987 were analysed for evidence of publication bias. They found that 52% of the 285 studies had been published and studies with statistically significant results ($p<0.05$) were more likely to be published than those showing a non-significant trend ($p>=0.05$) or no difference between the study groups. Dickersin and colleagues (1992) also investigated the association between publication and the significance of the results. A total of 737 studies approved by the two institutional review boards, which served the Johns Hopkins Health Institute in Baltimore, USA before 1981 were analysed. Information was obtained by contacting the principal investigator of the studies. They found evidence of publication bias for both institutional review boards, with a statistically significant association between significance of a study's result and publication.

Dickersin and colleagues also found that projects associated with the work of doctoral and master theses were less likely to be published than other types of projects.

A similar study investigated the association between publication and significance of the result in clinical trials funded by the United States National Institutes of Health in 1979 (Dickersin and Min, 1993). Of the 198 clinical trials completed by 1988, 93% had been published and trials with significant results were more likely to be published than those showing non-significant results. Most trials remained unpublished because the investigators thought that the results were "not interesting enough" or because they "did not have enough time". Bardy (1998) assessed the extent of possible selective reporting of the results of clinical trials in a study of 274 clinical drug trials that were notified to the Finnish National Agency for Medicines in 1987. By December 1993, final reports had been received by the agency for 38% of trials with positive results, 18% of trials with inconclusive results and 20% of trials with negative results; the outcomes for the remaining trials could not be assessed. The results of this study support the evidence of selective reporting, with trials with positive outcomes more likely to submit a final report to a regulatory authority than those with inconclusive or negative outcomes.

There is also evidence to show that full publication of research findings initially presented as conference abstracts is dependent on the statistical significance of their results. Scherer and Langenberg (2004), in a Cochrane methodology review of 46 research studies found that only 45% of abstracts initially presented as conference

abstracts were subsequently published in full. This figure increased slightly to 56% for those abstracts, which only presented the results of randomized trials. Most studies were published in full within three years of presentation at the meeting. Trials with statistically significant results were more likely to be published than those with non-significant or null results. This is important as results initially presented at scientific meetings are usually summarized as abstracts and published in conference proceedings or in journal supplements, and thus is not generally available. This may result in publication bias as the decision for publication is based on the magnitude or direction of study results. These findings are also consistent in areas of research other than healthcare. For example, a study of conference abstracts for research into the methods of health care evaluation, found that approximately half of the abstracts had not been, or were unlikely ever to be, published in full (Hopewell and Clarke, 2001).

The most common reason given by investigators as to why research submitted to scientific meetings is not subsequently published in full was lack of time. Other reasons given included that the researchers thought that a journal was unlikely to accept their study, or because the authors themselves perceived that the results were not important enough (Callaham et al, 1998; Weber et al, 1998; Donaldson and Cresswell, 1996; Eastwood et al, 1996).

Time-lag bias

Not only is publication dependent on the significance of the study results, there is also evidence to show that a 'time-lag' bias exists whereby research studies with significant

results are more likely than those with negative or null results to be stopped earlier than originally planned, published quicker or both. This may have important implications for the timing of the initiation and updating of a review, especially if there is an association between the inclusion of a trial in a review and its publication status.

A recent systematic review aimed to determine the extent to which the time to publication of a clinical trial was influenced by the significance of its results (Hopewell et al, 2004b). Studies were identified by searching the *Cochrane Methodology Register*, MEDLINE, EMBASE, Science Citation Index and by searching by hand journals and conference abstracts. Two studies met the inclusion criteria (Stern and Simes, 1997; Ioannidis, 1998). In both studies just over half of all trials had been published in full. Trials with positive results (i.e. those with statistically significant results ($p < 0.05$) in favour of the experimental arm of the trial) tended to be published in approximately four to five years. This was quicker than for those with null or negative results (i.e. those with results that were not statistically significant ($p = > 0.05$) or were statistically significant ($p < 0.05$) in favour of the control arm of the trial) with a time to publication of around six to eight years. One of the studies suggests that this difference could, in part, be attributed to the length of time taken to publish the results of a trial once follow up has been completed. This study showed that trials with null or negative findings took, on average, just over a year longer to be published than those with positive results.

Language bias

Excluding trials reported in languages other than English from a systematic review may also introduce bias (Moher et al, 2003; Jüni et al, 2002; Egger et al, 1997; Moher et al, 1996). Jüni and colleagues (2002) examined 303 meta-analyses and found that estimates of treatment effect were more beneficial in non-English language trials than in English language trials. However, the methodological quality of non-English trials tended to be lower than trials published in English, suggesting that comprehensive literature searches followed by careful assessment of trial quality are required to assess the contribution of all relevant trials, irrespective of language of publication.

Citation bias

Those carrying out systematic reviews often search the reference lists of articles to identify additional studies that may be relevant for inclusion in their review. The problem with this approach is that the articles retrieved in this way may be a biased sample and not representative of the totality of the evidence. The more a study is cited the easier it is to identify it for inclusion in a systematic review. Gøtzsche (1987) analysed the citations of trials published in articles assessing the use of non-steroidal anti-inflammatory drugs in rheumatoid arthritis. He found that trials demonstrating a positive effect of the new drug were more likely to be cited than trials showing a negative result. Similarly, trials for cholesterol lowering to prevent coronary heart disease were cited almost six times more often if they were supportive of cholesterol lowering, compared to unsupportive trials (Ravnskov, 1992). Citation bias may be a particular problem for those studies that remain unpublished or are cited in the grey

literature, as, for these types of studies, it will be even more difficult to identify an unbiased sample.

Duplication bias

The production of multiple publications from a single study can also lead to the introduction of bias. Studies with significant results are more likely to lead to multiple publications and presentations and therefore are more likely to be identified and included in a systematic review. Unless multiple publications of the same trial are clearly identified, the inclusion of duplicate data from these trials may lead to an overestimation of treatment effects (Tramer et al, 1997).

Bias and methodological quality

There is also evidence to show that the quality of the data included in a systematic review and/or meta-analysis can affect the overall results of the review. A study by Schulz and colleagues (1995) assessed the methodological quality of 250 controlled trials from 33 meta-analyses from the *Cochrane Pregnancy and Childbirth Database*. They found that trials with inadequate or unclear concealment of treatment allocation yielded larger estimates of treatment effect than trials reporting adequate concealment of treatment allocation. In addition, trials which were not double-blind, also yielded larger estimates of treatment effect.

A similar study (Moher et al, 1998) assessed the quality of 11 meta-analyses on the efficacy of interventions in circulatory and digestive diseases, mental health and

pregnancy and childbirth. This showed that trials with inadequate or unclear concealment of allocation were associated with an increased estimate of treatment benefit of 34% when compared to trials reporting adequate concealment of allocation. In a further study, Kjaergard and colleagues (2001) explored whether methodological quality affected estimates of treatment effect when comparing large and small randomized trials in meta-analyses. Fourteen meta-analyses containing 190 randomized trials were identified by searching *The Cochrane Library* and MEDLINE. They found that intervention effects were exaggerated in smaller trials reporting inadequate generation of the allocation sequence, inadequate concealment of allocation and inadequate blinding, when compared to large trials.

It is hoped that the CONSORT (Consolidated Standards of Reporting Trials) statement, which is designed to help authors improve the quality of reporting of randomized trials, will help improve the methodological quality of randomized trials (Moher et al, 2001b). The CONSORT statement advocates the use of a 22 item checklist pertaining to the title, abstract, introduction, methods, results and discussion. CONSORT also advocates the use of a flow diagram, which depicts information from four stages of a trial (enrolment, intervention allocation, followup and analysis) and includes the number of participants for each intervention group at each stage of the analysis.

Minimising bias through a comprehensive search

The identification of as unbiased and complete a set of relevant studies as possible for inclusion in a systematic review is therefore an important part of the systematic review process. The Cochrane Collaboration has done a considerable amount of work to identify reports of randomized trials and thereby minimise the risk of introducing bias into a systematic review. The *Cochrane Central Register of Controlled Trials* (CENTRAL) is now the most comprehensive source of records relating to controlled trials of healthcare interventions (Dickersin et al, 2002). As of July 2004, this register contains over 400,000 citations to reports of randomized trials and other studies potentially relevant for inclusion in systematic reviews of healthcare interventions.

CENTRAL includes reports of randomized trials that are not indexed in electronic databases such as MEDLINE or EMBASE, as well as those that are; it includes citations published in languages other than English; citations published only in conference proceedings and citations from other sources which are difficult to locate (Dickersin et al, 2002). Over 2,200 journals as well as hundreds of conference proceedings have been, or are being, handsearched (i.e. read from cover to cover) within The Cochrane Collaboration to identify reports of controlled trials to be included in CENTRAL (www.cochrane.org/Cochrane/hsearch.htm). In addition, a highly sensitive search strategy has been developed to identify reports of controlled trials in MEDLINE so that these records can also be included in CENTRAL (Dickersin et al, 1994). A similar project is currently underway to identify reports of randomized trials published in EMBASE and to include these in CENTRAL. As of July 2004,

nearly 70,000 relevant reports have been identified in EMBASE and submitted to CENTRAL (Anon, 2004).

The Campbell Collaboration is developing a similar register of reports of randomized trials of social, psychological, educational and criminological interventions. The register, which is called C2-SPECTR, currently contains approximately 12,000 reports of randomized or possibly randomized trials and serves as an important resource for identifying studies for inclusion in The Campbell Collaboration's systematic reviews. Methods similar to those used by The Cochrane Collaboration have been deployed to identify trials, including searching bibliographic databases such as the Educational Resources Information Center (ERIC), PsycINFO, Sociological Abstracts and Criminal Justice Abstracts and the handsearching of journals and conference proceedings (Petrosino et al, 2000).

If the search for reports of randomized trials is restricted to electronic databases, this has the potential to introduce bias into a review. A recent systematic review of studies compared the results of handsearching with those of searching electronic databases such as MEDLINE and EMBASE to identify reports of randomized trials. Handsearching identified 92-100% of the total number of reports of randomized trials found. The Cochrane Highly Sensitive Search Strategy, which is used to identify reports of randomized trials in MEDLINE, (Dickersin et al, 1994) identified 80% of the total number of reports of trials found. Electronic searches categorised as 'complex' found 65% and those categorised as 'simple' found 42%. The retrieval by an electronic search

was higher when the search was restricted to English language journals: 62% versus 39% for journals published in other languages. A major reason for the failure of the electronic searches to identify all the reports of randomized trials was because reports were published as abstracts and/or in journal supplements, which were not routinely indexed by the electronic databases. When the search was restricted to full reports, the retrieval for complex searches improved to 82%. Another reason for failure of the electronic searches to identify all the reports of randomized trials was because some reports were published before 1991 when the United States National Library of Medicine's system for indexing trial reports was not as well developed as it is today (Hopewell et al, 2004a).

Some studies will never be published in any format and, given the evidence that unpublished studies are likely to have different results to published studies, it is important to identify and include them in the review process to minimise bias (Dickersin, 1997). Contacting researchers, subject specialists and, in the case of health care, the pharmaceutical industry, can be an important means of identifying ongoing or unpublished studies. This is, however, dependent on the goodwill of the researchers and the time they have available to respond to these requests. Evidence suggests that requests for additional information about a study are not always successful (www.mrc-bsu.cam.ac.uk/firstcontact/pubs.html). To overcome this problem, some pharmaceutical companies are including and making publicly available, records of their ongoing and completed trials, either through the *meta*Register of Controlled Trials (www.controlled-trials.com) or by providing registers of research on their own web site (For example: www.schering.co.uk/). However, it is important to note that the

knowledge of a possible study does not necessarily mean that the data required for a review will be made available.

Research registers are also an important way of identifying ongoing research, some of which may never be formally published. There are many national registers such as the National Research Register in the UK, which contains ongoing health research of interest to, and funded by, the UK National Health Service (www.doh.gov.uk/research/nrr.htm). There are also numerous local and specialist registers each of which are specific to their own area of interest. Identifying ongoing studies is important so that whenever a review is updated, these studies can be assessed for possible inclusion. The identification of ongoing studies also helps the reviewer to assess the possibility of time-lag bias.

The Internet may be another important source of information about ongoing and completed research, particularly that which has not been formally published. However, searching the Internet can be a major undertaking. Many of the general search engines do not allow sophisticated searching and may identify thousands of web sites to check. Few studies have been conducted to establish whether searches on the World Wide Web (www) are a useful means of identifying additional unpublished and ongoing trials. One such study is by Eysenbach and colleagues (2001) who adapted search strategies from seven systematic reviews in an attempt to find additional reports of trials. They reviewed 429 web pages in 21 hours and found information about 14

unpublished, ongoing or recently published trials. At least nine of these were considered relevant for inclusion in four of the seven systematic reviews.

Differences between health and social sciences

The potential for introducing bias when conducting a systematic review may be even greater in areas outside health care, such as the social sciences. However, the empirical evidence in this area is limited and is still very much in its infancy. The medical literature is predominantly published in peer-reviewed journals, some of which are accessible through large, sophisticated, bibliographic databases such as MEDLINE and EMBASE. In comparison, the social sciences literature is much more diverse. Grayson and Gomersall (unpublished) advocate that one of the major differences between health and social sciences is in the publication of research. Social sciences research is not predominantly published in peer-reviewed journals and is often published in books, governmental and organisational reports, and other sources that are not as widely available. Some research is indexed in bibliographic databases such as Sociological Abstracts, Social Sciences Citation Index and the Applied Social Sciences Index and Abstracts (ASSIA) database, but the coverage and quality of these databases vary greatly. There is also a lack of specific indexing terms available in many of these databases to identify research studies.

What is grey literature?

The inclusion of grey literature in systematic reviews can help to overcome some of the problems of publication bias, which can arise due to selective availability of data. The search for, and inclusion of, grey literature therefore forms an important part of the

systematic review process. The definition of what constitutes grey literature, however, varies and the terminology can often be confusing (McAuley et al, 2000; Song et al, 2000; Cook et al, 1993; Alberani et al, 1990; Van Loo, 1985; British Library; STN International). Grey literature is generally assumed to include literature that has not been formally published, has limited distribution or is not available through conventional channels (Auger, 1998). Examples of grey literature include conference abstracts, research reports, book chapters, unpublished data, dissertations, policy documents and personal correspondence.

The emergence of the idea of grey literature is not new and the concept of grey literature is beautifully illustrated in a quotation by Minde-Pouet in 1920, cited by Schmidmaier (1986):

"No librarian who takes his job seriously can today deny that careful attention has also to be paid to the 'little literature' and the numerous publications not available in normal bookshops, if one hopes to avoid seriously damaging science by neglecting these"

Acceptance of the term "grey literature" dates back to 1978. A seminar in York in the UK was recognised as a milestone for grey literature as a primary information source and resulted in the creation of the System for Information on Grey Literature in Europe database (STN International), which is now managed by the European Association for Grey Literature Exploitation (EAGLE). In November 1997, the term grey literature

was redefined at the Third International Conference on Grey Literature (Auger, 1998). The revised definition is widely accepted and is known as "The Luxemburg Convention" after the city in which the conference was held. They define grey literature as:

"*that which is produced on all levels of governmental, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers*"

Sources of grey literature

There are a number of electronic resources available which are relatively easily accessible and which may prove useful in searching for grey literature. Four key types of these electronic resources are: grey literature databases, which only include studies found in the grey literature; general bibliographic databases that contain studies published in the grey and conventional literature; specific bibliographic databases, which contain studies published in the grey and conventional literature in a specific subject area; and ongoing trial registers that contain information about planned and ongoing studies. Examples of some key sources in each of these categories are given in Box 1.

Box 1: Some key electronic sources to identify grey literature

The Cochrane Central Register of Controlled Trials (CENTRAL) –includes references to trial reports in conference abstracts, handsearched journals, and other resources (www.cochrane.org).

The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) – includes references to trials reported in abstracts and handsearched journals (www.campbellcollaboration.org/).

Grey literature databases
BiomedCentral - contains meetings abstracts and protocols of ongoing trials (www.biomedcentral.com/).

British Library's Inside Web - includes references to papers presented at over 100,000 conference proceedings (www.bl.uk/services/current/inside.html).

Dissertation Abstracts Online - includes US dissertations since 1861 and British dissertations since 1988 (www.lib.umi.com/dissertations/).

System for Information on Grey Literature in Europe (SIGLE) - includes research and technical reports, preprints, working papers, conference papers, dissertations and government reports (www.cas.org/ONLINE/DBSS/sigless.html).

General databases which contain grey literature
Biological Abstracts - indexes papers presented at over 1,500 meetings, symposia and workshops world wide (www.biosis.org/products_services/ba.html).

Science Citation Index - indexes meeting abstracts not covered in MEDLINE or EMBASE relevant to sciences and technology (www.isinet.com/isi/products/citation/sci).

Subject specific databases which contain grey literature
CINAHL - includes nursing theses, government reports, guidelines, newsletters, and other research relevant to nursing and allied health (www.cinahl.com/).

ERIC (Education Resources Information Center) - includes report literature, dissertations, conference proceedings, research analyses, translations of research reports relevant to education and sociology (www.askeric.org/Eric/).

PsycINFO - includes dissertations, book chapters and academic/government reports relevant to psychology and sociology (www.apa.org/psycinfo/).

Sociological Abstracts - covers social and behavioural sciences and includes dissertations, conference proceedings, book chapters from 1963 to date (www.csa.com/csa/factsheets/socioabs.shtml).

Ongoing trials registers
BiomedCentral - publishes trials together with their International Standard Randomized Controlled Trial Number (ISRCTN) and publishes protocols of trials (www.biomedcentral.com/clinicaltrials/).

ClincalTrials.gov - covers trials in the US of life-threatening diseases (http://clinicaltrials.gov).

Current Controlled Trials *meta*Register of Controlled Trials – contains records from multiple trial registers including the UK National Research Register, the Medical Editors' Trials Amnesty, the UK MRC Trials Register, and links to other ongoing trials registers (www.controlled-trials.com).

The Lancet (Protocol Reviews) – contains protocols of randomized trials (www.thelancet.com).

Use of grey literature in systematic reviews

Estimates of the use and type of grey literature in systematic reviews vary. Clarke and Clarke (2000) conducted a study of the references included in Cochrane protocols and reviews published in *The Cochrane Library* in 1999. Of the references to studies included in the reviews, 92% were references to journal articles, 4% were conference proceedings that had not been published in journals, 2% were unpublished material (for example personal communication, in press documents and data on file), and 2% were books and book chapters.

Mallett and colleagues (2002) looked at the sources of grey literature included in the first 1,000 Cochrane systematic reviews published in Issue 1, 2001 of *The Cochrane Library*. Nine hundred and eighty-eight reviews were analysed; 12 reviews having been excluded, as the text of the review had been withdrawn from publication. All sources of information, other than journal publications, were counted as grey literature sources and were identified from the reference lists and details provided within the review. The "grey information" that was included ranged from details of study design, which helped to determine its eligibility for the review, to unpublished patient data included in the review analysis. The 988 reviews contained 9,723 studies in the "Included Studies" section of the review. Fifty-one per cent (n=513) of Cochrane reviews included trials with information from the grey literature. Table 1.2 shows the number of trial citations for different sources of grey information. The majority (55%) of the grey literature was from unpublished information supplementing published journal articles reporting trials. A random sample of 94 of these Cochrane reviews was

examined in detail to determine the nature of this unpublished information and it was found that 80% of the reviews listed the trial authors as the source of the unpublished information, with the remaining 20% of reviews not indicating the source of the unpublished information. The high proportion of unpublished information in Cochrane reviews contrasts with other studies where the major source of "grey information" was from conference abstracts (Egger et al, 2003; McAuley et al, 2000). This is possibly because Cochrane reviewers are specifically encouraged to search for and record unpublished information.

Table 1:2: Citations for grey literature in 513 Cochrane reviews

Grey literature source	**Number of trial references***
Unpublished information	1,259 (55%)
Conference abstracts	805 (35%)
Government reports	78 (4%)
Company reports	66 (3%)
Theses/dissertations	63 (3%)
Total grey literature citations (in 1,446 trials)	2,271 (100%)

*Trials can be referenced by more than one grey literature source. 4,820/6,266 trials were referenced only by published journal articles. 17 trials had missing citations.

Within reviews containing grey literature, studies containing some form of grey literature contributed a median of 28% of the included studies (IQR 13-50%). In 40 reviews more than 95% of studies contained information from the grey literature and in 67 reviews more than 75% of studies contained information from the grey literature.

Nine per cent of the trials (588/6,266) included in the 513 Cochrane reviews using grey literature, were referenced only by grey literature sources and not by journal articles.

These trials were analysed to determine how many of them would be identified if different grey literature sources were searched. The trials were categorised by a primary reference assigned according to the following hierarchy going from the least to the most "grey": conference abstracts, government reports, company reports, theses/dissertations, unpublished source. Figure 1.2 shows that 68% of trials were referenced by conference abstracts, with primary references for 8% being government reports, 7% being company reports and 4% being theses or dissertations. The remaining 19% of trials were referenced from unpublished reference sources alone (for example personal communication with trialists, in press documents and data on file). A previous analysis of 102 trials referenced from grey literature sources in 38 meta-analyses found 62% referenced by conference abstracts and 17% from unpublished sources, with similar proportions to the Mallett and colleagues study coming from government and company reports and theses (McAuley et al, 2000). A study of 153 trials in 60 meta-analyses by Egger and colleagues (2003) found 45% from conference abstracts, 14% from book chapters, 3% from theses and 37% from other grey literature sources including unpublished, company and government reports. Overall, the results are similar across the studies. A lower overall percentage of trials referenced only from grey literature sources were found in the study by Mallett and colleagues: 9% compared to 22% in the study by McAuley et al (2000) and 20% in the study by Egger et al (2003). This could be due to different inclusion criteria for the meta-analyses within the three studies and also to The Cochrane Collaboration's policy of encouraging reviewers to update reviews regularly, which might lead to grey literature references being replaced as published journal articles become available.

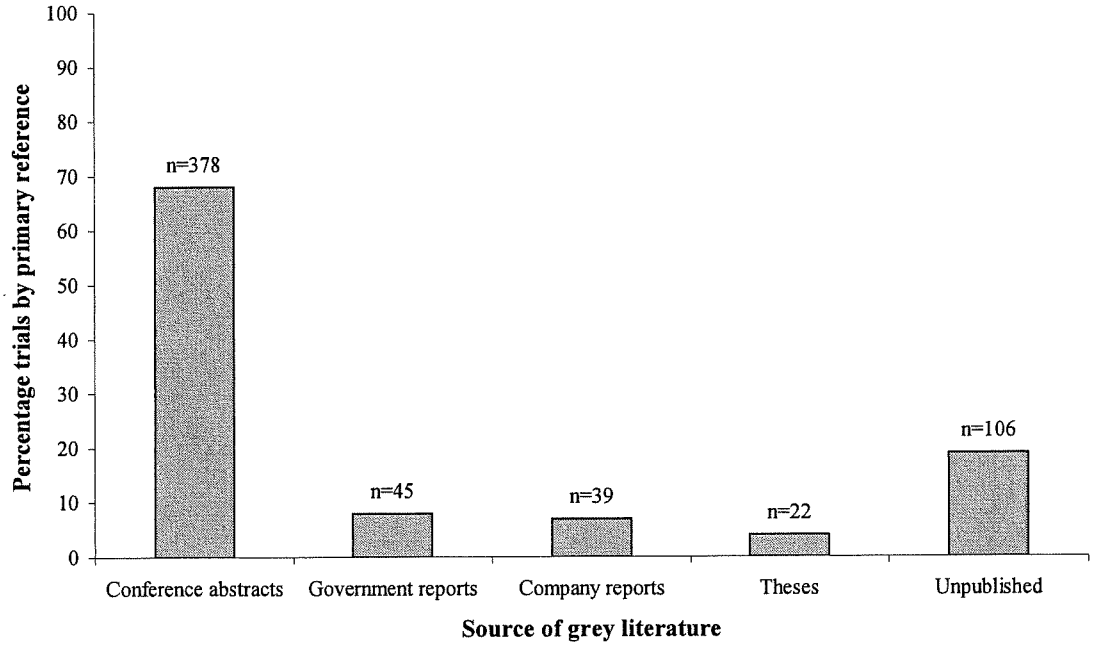


Figure 1:2: Source of primary reference for trials cited by grey literature alone included in Cochrane reviews in 2001

Further evidence of the use of grey literature in systematic reviews is demonstrated in a study assessing the source of references in 75 reviews of interventions for neurological conditions. Seventy-nine percent of references were from articles published in journals indexed in MEDLINE or EMBASE, 10% were from articles published in journals not indexed by these databases and 11% were from other sources such as books, pharmaceutical companies, data on file and unpublished studies (Taus et al, 1999).

Estimates of the amount of grey literature included in systematic reviews vary depending on the area of health care. Alberani and colleagues (1990) examined the type of references cited in articles published in six Italian health science journals in

1987 and 1988. A total of 1,398 articles were identified, 67% of the articles cited references found in the grey literature. Sixteen per cent of the total number of references identified were found in the grey literature. Of these, 69% were references to reports, 16% were references to conference proceedings, 11% were references to theses and 5% were references to other types of grey literature (these included unpublished data, personal communication or translations).

The use of grey literature in the social sciences and educational research may be even higher, although empirical evidence in this area is somewhat limited. In a review of the medical and psychological effects of unemployment, 72% of studies were published through conventional channels and 28% were found in the grey literature (Van Loo, 1985). In a systematic review by Petrosino and colleagues (2000) of randomized experiments in crime reduction, around one-third of the 150 trials identified were reported in government documents, dissertations and technical reports. Earlier, Boruch and colleagues (1978), in a review of early delinquency experiments, reported similar results in that about 25% of the included studies were unpublished.

Impact of grey literature in social sciences

Much of the early work on assessing grey literature and how it might affect the overall results of a systematic review was in the social sciences. In social sciences research, it has been known for some time that omitting unpublished studies from a systematic review can magnify the effect of the intervention. Glass and colleagues (1981) analysed 11 meta-analyses of the experimental literature published between 1976 and

1980. These included meta-analyses of psychotherapy, the effects of television on anti-social behaviour and sex bias in counselling. In each of the nine instances in which a comparison could be made, the average experimental effects from studies published in journals was larger than the corresponding effects estimated from theses and dissertations. The findings reported in journals were on average 33% more favourable to the hypothesis of the investigators than the findings reported in theses or dissertations. In one meta-analysis of sex bias in counselling and psychotherapy, the magnitude and the direction of the effect were different in the published and unpublished studies. A further study assessing socio-economic status and academic achievement, also found that the treatment effects reported in journals were greater than the corresponding effects reported in dissertations (White, 1982).

It is perhaps not surprising, given what is known about publication bias, that those studies that are most accessible, are most likely to be included in systematic reviews and are more likely to show larger treatment effects. A study of meta-analyses of psychological, educational and behavioural treatment research found that published studies were more likely to produce more positive results than unpublished studies. In the 92 meta-analyses studied, published studies yielded mean effect sizes that were, on average, 0.14 standard deviations larger than unpublished studies (Lipsey and Wilson, 1993). De Smidt and Gorey (1997) compared the results of a meta-analysis containing published studies, on the evaluation of social work interventions, with those of a replicated meta-analysis containing only studies found in the grey literature. The grey literature studies were doctoral dissertations and master theses, which reported

empirical findings on social interventions. Published studies showed the interventions to be more effective than studies found in the grey literature, however this difference was only marginal. The meta-analysis containing only published studies found that 78% of clients in the experimental group did better than those in the comparison group, compared to 73% for the meta-analysis containing only theses and dissertations.

Impact of grey literature in health sciences

There is now some evidence in health research, which shows that there is a systematic difference between the results of published studies compared to those found in the grey literature. An early example is a meta-analysis of preoperative parental nutrition for reducing complications from major surgery and fatalities (Detsky et al, 1987). In this study, the results of randomized trials published only as abstracts found the treatment to be more effective than trials published in papers; this is the opposite result to what might have been expected.

Several studies have suggested that the exclusion of grey literature from meta-analyses can lead to an exaggeration of the effect of treatment (Egger et al, 2003; McAuley et al, 2000). McAuley and colleagues (2000) examined whether the inclusion and exclusion of grey literature, in 41 separate meta-analyses, provided different estimates of treatment effectiveness. They found that grey literature accounted for between 5% and 75% of the studies in the meta-analyses and that, on average, published studies yielded significantly larger estimates of the intervention effect by 15% compared to grey literature studies. A similar study by Egger and colleagues (2003) assessed the impact

of the inclusion and exclusion of grey literature on the overall results of 60 meta-analyses. The meta-analyses were identified by handsearching journals, searching the *Cochrane Database of Systematic Reviews* and by searching the *Database of Abstracts of Reviews of Effects*. Only meta-analyses that contained more than five trials and had used comprehensive searches were eligible for inclusion. In this study, grey literature contributed to 14% of the total number of trials included in the meta-analyses and published trials yielded 7% larger estimates of the intervention effect compared to grey literature trials.

Problems in searching for grey literature

The identification of relevant studies in the grey literature and their inclusion in systematic reviews can be particularly time-consuming and difficult. In an attempt to determine whether researchers search for, obtain and then include grey literature in meta-analyses, Cook and colleagues (1993) identified all articles indexed with the keyword 'meta-analysis' in MEDLINE from January 1989 to February 1991. One hundred and fifty meta-analyses were identified; of these 33% did not search for unpublished information, 53% searched for unpublished information and 31% of these included unpublished information in the primary analysis. It was unclear whether unpublished information had been searched for in the remaining 13% of the meta-analyses.

The Cochrane Collaboration recommends that those carrying out systematic reviews take practical steps to avoid publication bias by trying to identify and include data from

trials found in the grey literature. To see how this recommendation was implemented in practice, Stewart and Tierney (2000) assessed all new reviews published in Issue 1, 2000 of *The Cochrane Library* to see whether Cochrane reviewers searched for grey literature trials and unpublished trials and whether that information was included in the reviews. They found that 71% of reviews reported searching sources other than bibliographic databases, 75% of reviews reported communicating with trialists and 52% specifically mentioned seeking unpublished trials. Of the 52 new reviews assessed, 59% included only published trials, 29% included trials reported as abstracts and letters, and 20% included unpublished trials.

Even when researchers do search for grey literature, there is some evidence to suggest that the considerable time, effort and expense may not always be worthwhile. In a massive attempt to obtain information about unpublished trials in perinatal medicine, Hetherington and colleagues (1989) sent letters to 42,000 obstetricians and paediatricians in 18 countries. They were notified of 395 unpublished randomized trials. Only 18 of these trials had been completed more than two years before the survey; 125 had ceased recruitment within the two years prior to the survey, 193 were actively recruiting at the time of the survey and 59 were about to begin recruitment. They concluded that obtaining information about trials retrospectively was unsuccessful but noted that the response rate to requests for details about ongoing and planned trials was good.

Difficulties of including grey literature

There is some debate as to whether studies found in the grey literature should be included in meta-analyses because they might be incomplete and their methodological quality can be difficult to assess. A survey by Cook and colleagues (1993) showed that 78% of meta-analysts and methodologists felt that unpublished material should definitely or probably be included in meta-analyses. Journal editors were less enthusiastic about the inclusion of unpublished studies; 47% of editors felt that unpublished data should probably or certainly be included in meta-analyses. Interestingly, Cook found that the editors' and methodologists' attitudes varied depending on the type of material included. Their attitudes were less favourable towards the inclusion of journal supplements, published abstracts, published symposia, book chapters and dissertations (none of which may have been peer reviewed) and more favourable towards unpublished presentations or materials that have never been published or presented. This study raised concerns that studies found in the grey literature may be of poorer quality, as they have not been peer reviewed. Traditionally, the peer review process has been thought to ensure the scientific quality of the studies it reviews and to help guarantee against scientific fraud. In reality, however, peer review still relies on the opinions of expert reviewers, which can be open to arbitration and bias (Jefferson et al, 2004).

There is some empirical evidence to show that studies found in the grey literature may be of poorer methodological quality than those published as journal articles. In an analysis of 60 meta-analyses of healthcare interventions, Egger and colleagues (2003)

found that trials published in the grey literature were less likely to report adequate concealment of allocation than published trials (34% versus 41%). Trials published in the grey literature were also less likely to report blinding of outcome assessment compared to published trials (45% versus 66%). The United States Food and Drug Administration (FDA) has a repository of unpublished data on the efficacy and safety of pharmaceuticals. MacLean and colleagues (2003) compared the quantity and quality of data obtained from the FDA with that of published data on the risk of non-steroidal anti-inflammatory drugs inducing dyspepsia in adults. There were significant differences in methodological quality between what was reported in the published studies compared to the FDA studies. The published studies were more likely to report appropriate methods of randomization (30% versus 8%), more likely to report appropriate methods of blinding (60% versus 3%) and more likely to report concealment of allocation (25% versus 0%). Withdrawals and dropouts were equally well reported in both the published literature and the FDA studies (85% versus 78%). MacLean et al found that the risk of developing dyspepsia was slightly greater in the published studies compared to the FDA studies, however in neither case was this difference significant.

Studies presented at scientific meetings and reported only as abstracts in the proceedings of the meeting are a principal source of grey literature. However, concerns have been raised over the poor reporting and quality of these abstracts. Several studies have highlighted that many abstracts submitted to scientific meetings may have had

inadequacies in the aims, methods, results and conclusions of the study or, indeed, have lacked numerical or statistical data (McIntosh, 1996; Panush et al, 1989).

Several studies have also highlighted some significant differences between the content and quality of information reported about a study in a conference abstract compared to its subsequent full publication (Bhandari et al, 2002; Chokkalingam et al, 1998; Weintraub, 1987). One study examined the consistency of information reported in an abstract and presented at an annual Orthopaedic surgeons conference in 1996, with that of its subsequent full publication. Bhandari and colleagues (2002) identified significant inconsistencies in the study title and authorship between what was reported in the abstract and its subsequent full publication. They also identified differences in the number of participants included in the full publication compared to what was reported in the abstract.

Further research is therefore needed to help assess the potential implications for those conducting systematic reviews of including, or not including, trials reported in the grey literature in their reviews and this thesis contributes to the research needed.

RESEARCH OUTLINE

Aims

The aims of this thesis are to examine the *actual* and *potential* impact of including trials reported in the grey literature in systematic reviews of randomized trials of healthcare interventions.

Objectives

This will be achieved through the following objectives:

- To systematically review empirical studies, which have investigated the impact of including grey literature in meta-analyses of randomized trials of healthcare interventions.

- To assess the actual impact of including grey literature in a large sample of meta-analyses of randomized trials. To assess the impact that different types of grey literature have on the meta-analyses and the importance of the methodological quality of published and grey literature trials.

- To assess the potential impact of retrospectively including trials reported in the grey literature in a large sample of systematic reviews.

- To assess in detail the content and methodological quality of trials reported in the grey literature and whether this has improved over time.

- To assess whether the content and methodological quality of the information reported for a trial in the grey literature differs from that which is subsequently published in full.

Research approach

This research was conducted using a staged approach. Each of the objectives described above was tackled as a specific piece of research. The findings from each of these pieces of research highlighted limitations in the current evidence, thereby forming the basis for the next objective and piece of research. Each objective forms a separate chapter of this thesis and, as such, details of the research methodology, results, discussion of the research findings and limitations of the existing evidence are described in each chapter.

The first objective described in chapter two, is a systematic review of current knowledge about including trials found in the grey literature in systematic reviews of healthcare interventions. Chapter three builds on these findings and aims to assess the impact of including trials reported in the grey literature in a large sample of meta-analyses drawn from reviews in cancer published in the *Cochrane Database of Systematic Reviews*. Chapter four, takes this a step further and tries to assess the potential impact of retrospectively including a large sample of trials found in the grey

literature in systematic reviews. In tackling this research question, a number of practical issues in trying to locate and include these trials in systematic reviews were identified, namely the quantity and quality of the information available in such reports. Chapter five builds on these findings by assessing in detail the quality of trial reports published in the grey literature, specifically those reported as abstracts in the proceedings of scientific meetings, and assesses whether this has changed over time. The final stage of this thesis, described in chapter six, assesses how the information reported for a trial in the grey literature differs from that which is eventually published in a full report of the trial. The concluding chapter integrates the findings from each of these research objectives and discusses the potential implications for reviewers of including or not including trials from the grey literature in systematic reviews. It also describes the contribution of this research to the current knowledge of the methodology of research synthesis. Recommendations for the preparation of systematic reviews and recommendations for further research are also discussed.

Throughout this thesis, including this background chapter, where work is described that has been carried out by other people, the examples used have been identified systematically and where possible represent the totality of the available evidence for the particular issue. Where a systematic review of the current evidence is available, this is provided. Where no systematic review was identified, a narrative discussion of the available evidence is described. Relevant material for this thesis was identified primarily from the *Cochrane Methodology Register,* which is published in *The Cochrane Library*. This register contains reports of empirical research relevant to the

methodology of systematic reviews of healthcare and social interventions, and other evaluations of health and social care. As of July 2004, it includes records for more than 5,000 published reports, conference abstracts, books, book chapters and unpublished material. Other relevant studies were identified by searching the *Cochrane Database of Methodology Reviews*, the UK Health Technology Assessment monographs, the U.S. National Library of Medicine's PubMed database, handsearching the proceedings of relevant conferences (such as the biennial Symposium on Systematic Reviews held in the UK and the annual international Cochrane Colloquia) and through personal contact with other researchers in the field.

CHAPTER 2: SYSTEMATIC REVIEW OF THE LITERATURE

Introduction

Chapter one of this thesis described how the inclusion of grey literature in systematic reviews may help to overcome some of the problems of publication bias, which can arise due to the selective availability of data. It showed that in social sciences research, it has been known for some time that omitting unpublished studies from a meta-analysis can magnify the effect of the intervention. For example, in a study of 11 meta-analyses published between 1976 and 1980, the average experimental effect found in studies published in journals was larger than the corresponding effect from theses and dissertations (Glass et al, 1981). There is now some evidence in support of this in health research, suggesting that the exclusion of grey literature from meta-analyses can lead to an exaggeration of the effect of treatment (Egger et al, 2003; McAuley et al, 2000).

This chapter aims to systematically review research studies that have investigated the impact of including trials reported in the grey literature in meta-analyses of healthcare interventions. A detailed version of this review has been published in the *Cochrane Database of Methodology Reviews* in *The Cochrane Library* (Hopewell et al, 2004c) (Appendix 1) and has been presented at the Fourth Symposium on Systematic Reviews in Oxford, UK, and the 10th International Cochrane Colloquium in Stavanger, Norway, in 2002 (Hopewell et al, 2002a; Hopewell et al, 2002b).

Objectives

To systematically review empirical studies, which have investigated the impact of including grey literature in meta-analyses of randomized trials of healthcare interventions.

Studies and data to be considered

Types of study

A research study was considered eligible for inclusion if it compared the impact of the inclusion and exclusion of grey literature on the results of one or more meta-analyses of randomized trials. The definition of what constitutes grey literature was that used by the authors in each of the empirical studies.

Data to be extracted from these studies

Information was collected from each study on: the area of health care, the number of meta-analyses, the number of randomized trials, the number of participants in these trials, the year of publication of the trials, the language of publication of the trials, the number and type of grey and published literature, and the methodological quality of the empirical studies and the randomized trials they included.

Main outcome measure

The main outcome measure was an estimate of the impact of trials from the grey literature on the pooled effect estimates of the meta-analyses. It had been hoped to do

this by calculating the ratio of the odds ratio between the results of the grey literature and the results of the published literature and estimating the percentage change. However, in reality, the summary outcome measures were frequently reported differently in each of the empirical studies. The summary outcome measure reported by the authors of the included studies was therefore used.

Search strategy for identification of studies

Studies were identified by searching the *Cochrane Methodology Register* (*The Cochrane Library* Issue 1, 2002) and MEDLINE (1966 to February 2002). The following search strategy was run against MEDLINE on OVID.

1 (Meta-Analysis or Clinical-Trials or Randomized-Controlled-Trials or Selection-Bias or Review-Literature).sh

2 (meta-analys$ or systematic review$ or data synthes$).tw

3 or/1-2

4 (grey literature or gray literature or unpublished).tw

5 exp Publication Bias/

6 or/4-5

7 3 and 6

Studies were also sought during the handsearching of selected journals, which is being carried out by the UK Cochrane Centre for all studies relevant to the methodology of systematic reviews. The abstracts presented at all Cochrane Colloquia (1993 to 2001),

Systematic Reviews Symposia (1998 to 2002) and the Society for Clinical Trials Meetings (1980 to 2001) (as published in *Controlled Clinical Trials*) were also handsearched as part of this activity.

The titles and abstracts of records retrieved with these strategies were assessed for relevance to the review. References in relevant reports were also checked to identify additional studies and the Science Citation Index (April 2002) was used to identify articles that cited relevant reports. Finally, researchers who may have carried out relevant studies were also contacted to ask if they knew of any additional studies. As of June 2002, searches had identified nearly 700 citations and full papers were obtained for 38 records.

Methods of the review

The titles and abstracts of all retrieved records were screened by one person (Sally Hopewell) to identify obvious exclusions (i.e. records that were clearly irrelevant to this review but were found by the electronic searches noted above). A second person (Steve McDonald) also checked less obvious records, before rejection. Full copies of the reports were obtained for each of the non-rejected records and these reports were assessed independently to determine if they met the inclusion criteria for the review. Any disagreements were resolved through discussion.

Two people (Sally Hopewell and Steve McDonald) independently extracted data from each of the included studies as described above. If any of the data for an empirical

study were insufficient or missing, these were sought from the named contact author (or the responsible person in the case of unpublished studies).

The decision on whether or not to combine the results of the included empirical studies was dependent on an assessment of heterogeneity. It had been hoped to carry out a meta-analysis of those studies that were sufficiently homogenous in their design. However, the summary outcome measures in the empirical studies were frequently reported differently and therefore could not be combined. The outcome measures reported by the authors of the empirical studies (for example: relative risk instead of odds ratio) have therefore been presented and analysed descriptively.

Results

Description of studies

Eight studies met the inclusion criteria for this review (Burdett et al, 2003; Egger et al, 2003; Horn and Limburg, 2002; Bhandari et al, 2000; Fergusson et al, 2000; McAuley et al, 2000; Man-Son-Hing et al, 1998; Jeng et al, 1995) and all were published as full papers. For one study, additional information was also obtained from the authors (Fergusson et al, 2000).

The included studies covered a variety of different areas of health care. Some studies covered specific areas of health care such as orthopaedic and cardiac surgery, cancer, stroke and miscarriage, while others included meta-analyses from a number of different medical specialities. The meta-analyses included in these studies were identified from a

variety of sources, including searches of MEDLINE and trial registers, handsearches of journals and by contacting researchers in the area of interest and the pharmaceutical industry. Within the included studies, the definition of what constituted grey literature varied from study to study. However, in general, it tended to include abstracts, book chapters, unpublished data, theses, company reports and letters. More information is available in Table 2.1.

Characteristics of the meta-analyses

Of the eight included studies, four (Burdett et al, 2003; Egger et al, 2003; Fergusson et al, 2000; McAuley et al, 2000) contained multiple meta-analyses, which compared the inclusion and exclusion of grey literature on the pooled effect estimate of the meta-analyses. One of these (Burdett et al, 2003) used individual patient data meta-analyses. The remaining four included studies (Horn and Limburg, 2002; Bhandari et 2000; Man-Son-Hing et al, 1998; Jeng et al, 1995) were single meta-analyses where the difference between the treatment effect of grey literature trials (grey trials) and published trials was explored in a sensitivity analysis. One of these (Man-Son-Hing et al, 1998) was an individual patient data meta-analysis. The two studies containing the greatest number of meta-analyses were the studies by Egger and colleagues (2003), which contained 60 meta-analyses, and the study by McAuley and colleagues (2000) that contained 41 meta-analyses (Table 2.2).

Table 2:1: Characteristics of the included studies

Study	Methods	Data	Comparisons	Outcome measures
Bhandari et al, 2000	Single meta-analysis carried out as a sensitivity analysis in a systematic review. Trials were identified by searching MEDLINE and SCI Search (1969-1998). Grey literature was defined as abstracts and unpublished manuscripts.	One meta-analysis for reamed versus nonreamed intramedullary nailing of lower long bone fractures, containing 9 RCTs and 714 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.	Published trials showed a greater treatment effect than grey trials; this difference was not statistically significant. Relative risk 0.43 (95% CI 0.19-0.99) for published trials versus 0.3 (95% CI 0.11-0.83) for grey trials.
Burdett et al, 2003	Multiple individual patient data meta-analyses. Trials were identified by electronic searches, handsearching, trial registers and by contacting trialists. Grey literature was defined as unpublished data, abstracts, book chapters and non-English language.	11 meta-analyses of cancer treatments, containing 120 RCTs and 18,377 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.	Published trials showed a greater treatment effect than grey trials; this difference was statistically significant. Ratio of hazard ratios 0.93 (95% CI 0.90-0.97) for published trials versus 0.96 (95% CI 0.93-0.99) for all trials.
Egger et al, 2003	Multiple meta-analyses were identified by handsearching eight high-impact medical journals (1994-1998), searching CDSR (Issue 1, 1998), DARE (1994-1998) and handsearching HTA reports. Meta-analyses were eligible if they combined binary outcomes of at least five trials and had comprehensive literature searches. Grey literature was defined as conference abstracts, books, theses, file drawer data and material from trial registers.	60 meta-analyses from a number of different medical specialties, containing 783 RCTs and 167,733 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.	Published trials showed a greater treatment effect than grey trials; this difference was not statistically significant. Ratio of odds ratios 1.07 (95% CI 0.98-1.15) for grey trials versus published trials.
Fergusson et al, 2000	Multiple meta-analyses identified from a project undertaken by the International Study of Peri-operative Transfusion. Trials were identified by searching MEDLINE and EMBASE (1966-1996). Grey literature was defined as abstracts, letters and conference proceedings.	10 meta-analyses of methods to reduce the need for peri-operative transfusion, containing 114 trials and 11,142 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.	Published trials showed a greater treatment effect than grey trials; this difference was not statistically significant.

Study	*Methods*	**Data**	**Comparisons**	**Outcome measures**
Horn and Limburg, 2002	Single meta-analysis carried out as a sensitivity analysis in a Cochrane review. Trials were identified by searching the Cochrane Stroke Review Group's specialised register. Grey literature was defined as abstracts and unpublished data.	One meta-analysis for calcium antagonists for acute ischaemic stroke, containing 21 RCTs and 6414 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.	Published trials showed no treatment effect. Grey trials showed a negative treatment effect. This difference was not statistically significant. Odds ratio 1.03 (95% CI 0.92-1.15) for published trials versus 1.33 (95% CI 1.00-1.79) for grey trials.
Jeng et al, 1995	Single meta-analysis carried out as a sensitivity analysis in an individual patient data meta-analysis. Trials were identified through the ethics committee of the American Society for Reproductive Immunology. Grey literature was defined as unpublished.	One meta-analysis of paternal immunisation for recurrent miscarriage, containing 8 trials and 379 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.	Published trials showed a greater treatment effect; this difference was not statistically significant. Relative risk 1.17 (95% CI 0.97-1.37) for published trials versus 1.01 (95% CI 0.74-1.28) for grey trials.
Man-Son-Hing et al, 1998	Single meta-analysis carried out as a sensitivity analysis in a systematic review using individual patient data. Trials were identified by searching MEDLINE and EMBASE (1994-1997). Grey literature was defined as data obtained from the pharmaceutical industry.	One meta-analysis of quinine for nocturnal leg cramps, containing 7 trials and 409 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.	Published trials showed a greater treatment effect; this difference was not statistically significant. Relative risk reduction 0.43 (95% CI 0.21-0.65) for published trials versus 0.15 (95% CI 0.06-0.24) for grey trials.
McAuley et al, 2000	Multiple meta-analyses were selected at random from an existing database of meta-analyses established by searching MEDLINE (1966-1995). Grey literature was defined as abstracts, unpublished studies, conference proceedings, theses, book chapters and company reports.	41 meta-analyses from a number of different medical specialities, containing 467 RCTs and 217,427 participants.	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.	Published trials showed a greater treatment effect; this difference was statistically significant. Ratio of odds ratio 1.15 (95 % CI 1.04-1.28) for grey versus published trials. Ratio of odds ratio 1.02 (95% CI 0.91-1.14) for abstracts versus published trials. Ratio of odds ratio 1.33 (95 % CI 1.10-1.60) for grey (excluding abstracts) versus published trials.

Table 2:2: Characteristics of the meta-analyses

Study	No. meta-analyses	No. trials	No. participants	No. grey trials	No. pub trials	Area health care	Year publication
Bhandari 2000	1	9	714	5	4	Orthopaedic surgery	1996-1998
Burdett 2003	11	120	18,377	45	75	Cancer	-
Egger 2003	60	783	167,733	153	630	Various medical specialities	1994-1998
Fergusson 2000	10	114	11,142	6	108	Cardiac and orthopaedic surgery	-
Horn 2002	1	21	6,414	4	17	Acute ischaemic stroke	1984-1999
Jeng 1995	1	8	379	4	4	Paternal immunisation for recurrent miscarriage	1986-1993
Man-Son-Hing 1998	1	7	409	3	4	Quinine for nocturnal leg cramps	1987-1993
McAuley 2000	41	467	217,427	102	365	Gastro 24% Cardiac 21% Infection 12% Reproduction 12% Circulatory 9% Other 21%	-

Characteristics of the trials included in the meta-analyses

The number of trials included in the meta-analyses ranged from seven to 783 trials. Again, the studies by Egger and colleagues (2003) (783 trials) and McAuley and colleagues (2000) (467 trials) contained the greatest number of trials. In most cases, there were more published trials included in the meta-analyses than grey trials (median 46 (IQR 4-300) versus median six (IQR 4-88)). Grey trials were defined as those trials, which had not yet been published as full journal articles. The number of participants included in the trials ranged from 379 (Jeng et al, 1995) to 217,427 participants (McAuley et al, 2000) (Table 2.3), with published trials generally being larger than the grey trials. It was not possible to combine information on the

number of participants included in each trial as this information was reported differently across the individual studies.

The studies by Egger and colleagues (2003) and McAuley and colleagues (2000) also assessed the statistical significance of the trial results included in the meta-analyses. Egger and colleagues found that published trials (30%) were more likely to have statistically significant results ($p<0.05$) than grey trials (19%). This difference was not apparent in the study by McAuley and colleagues, who found that 66% of published trials had statistically significant results compared to 68% of grey trials.

Table 2:3: Characteristics of the trials included in the meta-analyses

Study	No. trials	No. participants	No. participants per trial
Bhandari 2000	Published: 4 Grey: 5	Published: 366 Grey: 348	Published (median) 87.5 (IQR 57.7-129.2) Grey (median) 75 (IQR 49-87.5)
Burdett 2003	Published: 75 Grey: 45	Published: 12,156 Grey: 6,221	Published (mean)162 Grey (mean) 138
Egger 2003	Published: 630 Grey: 153	Published: 146,160 Grey: 21,573	Published (median) 102 (range 8-5042) Grey (median) 91 (range 9-1012)
Fergusson 2000	Published: 108 Grey: 6	Published: 10,322 Grey: 820	Published (median) 638 (IQR 240-828) Grey (median) 149 (IQR 85-250)
Horn 2002	Published: 17 Grey: 4	Published: 5,626 Grey: 788	Published (median) 189 (IQR 55-411) Grey (median) 219 (IQR 135-237)
Jeng 1995	Published: 4 Grey: 4	Published: 239 Grey: 140	Published (median) 54 (IQR 40-85) Grey (median) 33 (IQR 24-48)
Man-Son-Hing 1998	Published: 4 Grey: 3	Published: 73 Grey: 336	Published (median) 19 (IQR 10-26) Grey (median) 69 (IQR 62-205)
McAuley 2000	Published: 365 Grey: 102	Published: 23,286 Grey: 194,141	Published (median) 113 (IQR 59-228) Grey (median) 84 (IQR 48-190)

Language of publication of the trials

Information on the language of publication of the trials included in the meta-analyses was available for seven of the eight included studies. The majority of trials (96%) were published in English. This was similar for both grey and published trials and no language restrictions appear to have been imposed by the authors of the studies. Only three of the seven studies (Burdett et al, 2003; Fergusson et al, 2000; McAuley et al, 2000) contained trials published in languages other than English. These included trials published in French, German, Italian, Japanese and Spanish.

Type of grey literature

For all eight included studies the most common type of grey literature were abstracts (49%). Unpublished data were the second largest type of grey literature (33%). The definition of what constituted unpublished data varied since it used the definition used by the authors of the study, however, it included data from trial registers, file-drawer data and data from individual trialists. Book chapters were the third largest type of grey literature (10%), with unpublished reports, pharmaceutical company data, in press publications, letters and theses making up the small remainder (Figure 2.1). In one study (Man-Son-Hing et al, 1998) pharmaceutical company data was the only source of grey literature. This was an individual patient data meta-analysis where the authors of the study were in direct contact with the company who distributed the drug.

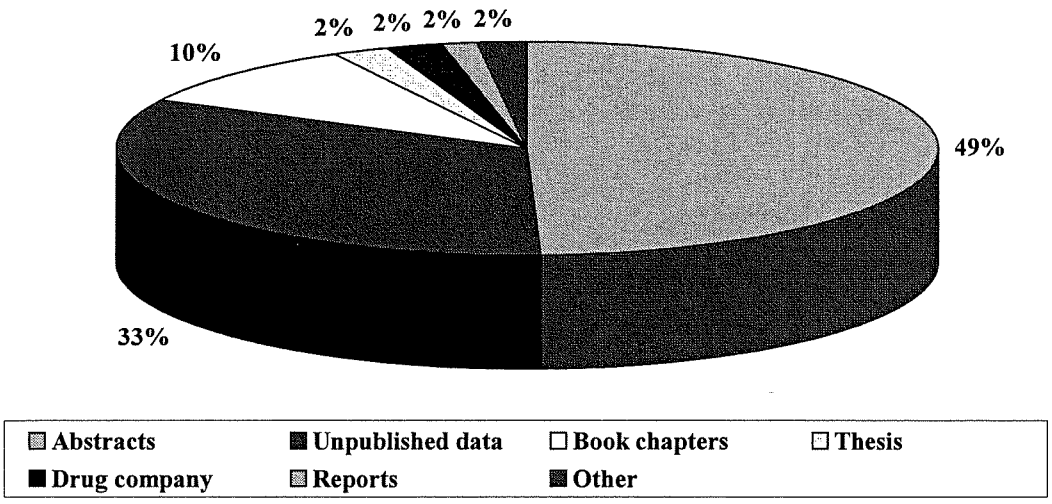


Figure 2:1: Type of grey literature included in the meta-analyses

Methodological quality of the randomized trials included in the meta-analyses

Information on the methodological quality of the trials included in the meta-analyses was only available for two of the eight included studies (Egger et al, 2003; Bhandari et al, 2000) (Table 2.4). Only one of these studies (Egger et al, 2003) contained multiple meta-analyses and they found that trials published in the grey literature were less likely to report adequate concealment of allocation compared to published trials (34% versus 41%). Grey trials were also less likely to report blinding of outcome assessment compared to published trials (45% versus 66%). However, in this study the quality of allocation concealment and blinding assessment were only assessed in the subset of trials that were included in Cochrane reviews. In the study by Egger and colleagues, the assessment of allocation concealment was based on the allocation concealment score (i.e. adequate, inadequate or unclear), which was assigned by the authors of the original systematic review. Egger and colleagues reported that sometimes this information was unclear

or had not been assessed by the authors of the reviews. However, they do not differentiate between the number of trials reporting inadequate methods of allocation concealment, the number of trials reporting unclear methods of allocation concealment and the number of trials where allocation concealment was not assessed by the authors of the original review.

Table 2:4: Methodological quality of the trials included in the meta-analyses

Study	No. trials	Concealment YES	Concealment NO/UNCLEAR	No. trials	Blinding YES	Blinding NO/UNCLEAR
Bhandari 2000	Grey 5 Published 4	Grey 2 (40%) Published 3 (75%)	Grey 3 (60% Published 1 (25%)	-	-	-
Egger 2003	Grey 77 Published 339	Grey 26 (34%) Published 138 (41%)	Grey 51 (66%) Published 201 (59%)	Grey 71 Published 345	Grey 32 (45%) Published 227 (66%)	Grey 39 (55%) Published 118 (34%)

Outcome measures for grey and published trials

The summary outcome measuring the impact of grey and published trials on the pooled effect estimate of the meta-analyses was reported differently across all eight included studies and, therefore, could not be combined formally. Two studies reported a ratio of odds ratios, one an odds ratio, one a relative risk, one a relative risk reduction, one study reported a ratio of hazard ratios and, for one study, data from the individual trials were not combined (Table 2.5).

Multiple meta-analyses

All four of the included studies containing multiple meta-analyses (Burdett et al, 2003; Egger et al, 2003; Fergusson et al, 2000; McAuley et al, 2000) found that published trials showed an overall greater treatment effect than grey trials (Table

2.5). This difference was statistically significant in two of these four studies (Burdett et al, 2003; McAuley et al, 2000). The study by McAuley and colleagues (2000) found that grey trials were likely to produce smaller estimates of the intervention effect, on average by 15%, compared with published trials (ratio of odds ratios 1.15; 95% CI 1.04-1.28).

Table 2:5: Summary outcome measures for grey and published trials

Study	Overall treatment effect	Grey trials	Published trials
Bhandari 2000	Relative risk 0.33 (95% CI 0.16-0.68)	Relative risk 0.30 (95% CI 0.11-0.83)	Relative risk 0.43 (95% CI 0.19-0.99)
Burdett 2003	Ratio of hazard ratios 1.04 (95% CI 1.01-1.08)	-	Ratio of hazard ratios 1.08 (95% CI 1.03-1.11)
Egger 2003	Ratio of odds ratios (grey versus published) 1.07 (95% CI 0.98-1.15)	-	-
Fergusson 2000	-	-	-
Horn 2002	Odds ratio 1.06 (95% CI 0.96-1.18)	Odds ratio 1.33 (95% CI 1.00-1.79)	Odds ratio 1.03 (95% CI 0.92-1.15)
Jeng 1995	-	Relative risk 1.01 (95% CI 0.74-1.28)	Relative risk 1.17 (95% CI 0.97-1.37)
Man-Son-Hing 1998	Relative risk reduction 0.21 (95% CI 0.12-0.30)	Relative risk reduction 0.15 (95% CI 0.06-0.24)	Relative risk reduction 0.43 (95% CI 0.21-0.65)
McAuley 2000	Ratio of odds ratios (grey versus published) 1.15 (95% CI 1.04-1.28)	-	-

In the study by Egger and colleagues (2003) the pooled estimates from grey trials were on average 7% less beneficial than estimates from published trials (ratio of odds ratios 1.07 95% CI 0.98-1.15). However, in the Egger study, there was notable heterogeneity between meta-analyses and a wide variation between different medical specialities. For example, in some areas there was little difference between the results of grey and published trials but in other areas, published trials showed a much larger treatment effect. In the review of individual patient data meta-analyses

by Burdett and colleagues (2003), published trials also showed a significantly greater treatment effect in favour of the experimental intervention (ratio of hazard ratios 1.08; 95% CI 1.03-1.011).

Single meta-analyses

Three of the studies containing single meta-analyses (Bhandari et al, 2000; Man-Son-Hing et al, 1998; Jeng et al, 1995) also found that published trials showed an overall greater treatment effect than grey trials, although this difference was not found to be statistically significant in any of the three studies. The remaining study, which was a single meta-analysis, found that published trials showed no effect of treatment and that grey trials showed a negative treatment effect (Horn and Limburg, 2002); again, this difference was not statistically significant (Table 2.5).

Only one of the eight studies (McAuley et al, 2000) assessed the type of grey literature and its impact on the overall results of the meta-analyses. McAuley and colleagues found that published trials showed an even greater treatment effect than grey trials if abstracts were excluded from the meta-analyses. The ratio of odds ratios for grey trials versus published trials was 1.15 (95% CI 1.04-1.28) compared to 1.33 (95% CI 1.10-1.60) for grey trials, excluding abstracts, versus published trials.

Discussion

This systematic review of the literature shows that published trials tend to be larger and may show an overall greater treatment effect, on average, than trials found in the grey literature. One study also showed that published trials are more likely to

have statistically significant results compared to grey trials. On average, there were more published trials included in the meta-analyses than grey trials. This may have particular implications for meta-analyses containing only a few trials where the impact of excluding trials found in the grey literature has the greatest potential to introduce bias.

Interestingly, the study by Egger and colleagues (2003), which contained the greatest number of meta-analyses, specifically excluded meta-analyses containing only a small number of trials. It is not clear if the study by McAuley and colleagues, which also contained a large number of meta-analyses, applied similar criteria. A recent assessment of systematic reviews published in *The Cochrane Library* found that the median number of trials included in a typical Cochrane review was six (Mallett and Clarke, 2002). If the Egger study had included meta-analyses containing only a small number of trials, it is anticipated that the difference in treatment effect, between published and grey trials, may have been greater.

The methodological quality of grey and published trials may also have an impact on the overall results of a systematic review. Only two of the included studies assessed the methodological quality of the trials included in the meta-analyses (the other six did not address this issue). One of these two studies contained multiple meta-analyses (Egger et al, 2003) and found that allocation concealment and blinding of outcome assessor were better in published trials than trials found in the grey literature. However, in the Egger study, the methods of allocation concealment and blinding were only assessed in the subset of meta-analyses that were Cochrane

reviews and sometimes the necessary information was unclear or had not been assessed by the authors of the original Cochrane review.

A recent study by Middleton (2003) documented how the methods of allocation concealment were described in a sample of Cochrane reviews. Middleton assessed 101 reviews (including 984 studies) published in Issue 1, 2003 of *The Cochrane Library*. She found that 39% (388/984) of studies were miscoded and, for 256 of these 388 studies (26% overall), reviewers failed to comment, at all, on the method of allocation concealment. Common reasons for miscoding were confusion of allocation concealment with the randomization process in general and confusion with the generation of the schedule and blinding of the intervention after randomization. This highlights possible problems with using the findings of the original reviewers to assess quality; in particular the problems of using the allocation concealment score assigned by authors of Cochrane reviews and may point to a need for a more independent assessment of trial quality.

The definition of grey literature varied across all eight included studies. Overall, abstracts were the most common source of grey literature. Only one of the included studies (McAuley et al, 2000) assessed the impact of the inclusion and exclusion of different types of grey literature on the overall results of the meta-analyses. The study by McAuley and colleagues (2000) found that the exclusion of abstracts from the meta-analyses showed that published trials had an even greater treatment effect than grey trials. It has been argued by some that trials reported in abstracts should not be included in meta-analyses as the data reported is often preliminary and may not be representative of the final results of the trial (Cook et al, 1993). However,

evidence suggests that just over half (56%) of all trials reported as abstracts and presented at conferences are subsequently published in full and that it takes on average, three years for a trial reported in an abstract and presented at a scientific meeting to be published in full (Scherer and Langenberg, 2004). This means that a large number of trials presented at scientific meetings will never be published in full. Further research is needed to understand better the effect of trials published in different types of grey literature on the overall results of systematic reviews.

The findings of this systematic review have important implications for people conducting systematic reviews who need to ensure they search for trials in the grey as well as published literature in order to minimise bias in their review. However, the findings of this review do not address the problems faced by those carrying out systematic reviews in the identification and retrieval of trials published in the grey literature; a process that can be particularly time-consuming and involve considerable cost and effort (Hetherington et al, 1989). In addition, if and when grey trials are identified their inclusion in a systematic review can sometimes be problematic as the relevant data may often be incomplete or missing (Cook et al, 1993).

A current limitation of this review is that it has not been possible to combine the summary outcome measures from the included studies, as they were frequently reported differently. The issue of how to combine the summary outcome measures is being explored. For example a meta meta-analysis approach, whereby the data from the individual trials included in the meta-analyses are combined, would help to overcome this problem and also allow greater exploration of issues such as the

quality of individual trials and the impact of different types of grey literature on the overall results of the meta-analyses.

A further limitation is that the studies that included a single meta-analysis were not specifically designed to address the impact of the inclusion and exclusion of grey literature on the overall results of the meta-analysis and had been carried out as a sensitivity analysis. These studies are subject to a possible outcome reporting bias as they may only have reported differences between grey and published trials if these were found.

Limitations of the existing evidence

The results of chapter two show that published trials tend to be larger and may show an overall greater treatment effect than those trials identified in the grey literature. Therefore, if only published trials are included in a systematic review then the results and conclusions of that review may be biased. However, this systematic review of the existing evidence has also highlighted a number of limitations:

- Further evidence is needed to understand better the effect of trials published in the grey literature on meta-analyses containing only a small number of trials.

- Further evidence is needed to understand better the effect of the quality of published and grey trials on the overall results of meta-analyses.

- Further evidence is needed to understand better the effect of trials published in different types of grey literature on the overall results of meta-analyses.

CHAPTER 3: ACTUAL IMPACT OF INCLUDING GREY LITERATURE IN META-ANALYSES OF RANDOMIZED TRIALS

Introduction

The evidence described in Chapter Two compares the impact of including and excluding trials found in the grey literature in systematic reviews of healthcare interventions. This looked at meta-analyses containing a relatively large number of trials. The average number of trials included in a Cochrane systematic review is six (Mallett and Clarke, 2002) and not all of these trials will necessarily be included in the statistical meta-analysis of the systematic review. There is, therefore, a need to understand the impact of trials found in the grey literature in meta-analyses containing a smaller number of trials, as this is where the exclusion of trials from the grey literature has the greatest potential to introduce bias. Current evidence also suggests that excluding trials reported as abstracts from a meta-analysis could mean that published trials show an even greater treatment effect than trials found in the grey literature (McAuley et al, 2000). In addition, it has also been suggested that trials found in the grey literature might be of poorer quality than published trials (Egger et al, 2003). Chapter three of this thesis aims to build on the existing evidence and attempts to address some of the areas where there is still uncertainty.

Objectives

To assess the actual impact of including grey literature in a large sample of meta-analyses of randomized trials of healthcare interventions. To assess the impact that

different types of grey literature have on the meta-analyses and to assess the importance of the methodological quality of published and grey trials on the meta-analyses.

Methods

Source of data

The sample of meta-analyses were taken from systematic reviews in cancer published in the *Cochrane Database of Systematic Reviews* in *The Cochrane Library*, Issue 3, 2003. Details were obtained from the Cochrane Cancer Network of all groups within The Cochrane Collaboration preparing and maintaining systematic reviews in the field of cancer. One hundred and thirty cancer reviews were identified from the 1745 systematic reviews published in the *Cochrane Database of Systematic Reviews* at that time.

Inclusion criteria

A systematic review was considered eligible for inclusion if it included one or more meta-analyses, one or more reports from the grey literature and one or more reports from the published literature.

Data extracted from each systematic review

The reference lists from each of the 130 Cochrane reviews were assessed and trials were classified as either published trials, grey trials or published trials supplemented with personal communication. The "Description of Studies" section and "Acknowledgments" section of each review were also assessed, in particular to

identify where published trials had been supplemented with personal communication.

Published trials were defined as those trials with reports published as full journal articles either in the regular issue of the journal or in a supplementary issue of the journal.

Grey trials were defined as those trial reports published as conference abstracts, letters, books, book chapters, government reports, pharmaceutical reports, theses or dissertations, and unpublished "file-drawer data" or personal communication with the trialists.

If it was unclear whether a reference to a trial was from a published or grey source then the U.S. National Library of Medicine's PubMed database (www.ncbi.nlm.nih.gov/entrez/query.fcgi) and / or *The Central Register of Controlled Trials* (CENTRAL) were searched for more information on the reference type. If it was still unclear, a copy of the report was obtained. If a trial had a published reference source and a grey reference source (for example an earlier report published as a conference abstract) then the published source was taken to be the primary reference source. The exception to this rule was if a published source was supplemented with personal communication from the authors of the trial; this was noted as a separate category.

Published trials supplemented with personal communication were defined as a published journal article where the authors of the review had also obtained

additional information from the authors of the trial. It was difficult to quantify the amount of additional information obtained from the authors of the trials as, for example, the information might have been to clarify the method of allocation concealment or the supply of additional data for an outcome which was not reported in the full publication.

Reviews containing individual patient data meta-analyses were excluded from the main analysis because these involved the central collection of raw data from each trial.

The majority of Cochrane reviews included randomized trials. However, a small number of Cochrane reviews included other studies, which were not randomized. These studies were not usually combined in a meta-analysis and were therefore excluded.

Data extracted from each meta-analysis

If a systematic review contained published and grey trials or published trials supplemented with personal communication in the "Included Studies" section of the review, the meta-analysis displaying the results of the primary outcome of the review (as defined by the authors of the review) was assessed to see if they were included. Where there was more than one primary outcome in a review the outcome with the greatest overall number of trials and the greatest number of grey trials was used. Only those meta-analyses displaying binary outcomes were assessed.

If the meta-analysis displaying the primary outcome contained only two trials or did not contain both published and grey trials then the second or third outcome with the greatest number of published and grey trials was analysed. Similarly, if the meta-analysis displaying the primary outcome was not pooled by the authors of the review because of excessive heterogeneity (as defined by the authors of the review) then the next available outcome with the greatest number of published and grey trials was used.

Data were extracted from each eligible meta-analysis on the area of cancer, the type of intervention, the type of comparison intervention, the type of primary outcome, and the number of published trials, grey trials and published trials supplemented with personal communication.

Data extracted from each trial

Data were extracted from each trial included in the meta-analysis on the date of the publication, the type of publication (e.g. abstract, letter, full paper, thesis), the size of the trial, the number of events and the number of participants in the intervention and comparison groups, the statistical model used by authors of the review (i.e. fixed effect or random effects model) and the type of effect estimate used (i.e. odds ratio, Peto odds ratio, relative risk or hazard ratio).

Data were also extracted on the quality of the trials included in the meta-analysis. Quality was assessed by checking the table of "Included Studies" section of the review and by reading the "Methodological Quality" section of the review.

Quality assessment was based on the allocation concealment score in the table of "Included Studies" where the authors of the review defined whether the method of allocation concealment was Adequate, Unclear or Inadequate. Authors were also able to select "Not Used" for this variable. Although it is recognised that there are potential problems with this approach for assessing trial quality, as it relies on the information reported by the authors of the systematic review (Middleton, 2003), this approach was deemed preferable to assessing the quality of each trial independently, as this would not have been feasible within the scope of the project.

Adequate allocation concealment is defined as when patients and investigators enrolling patients cannot foresee assignment (Schulz and Grimes, 2002a). Adequate methods included a priori numbered or coded drug containers of identical appearance, central randomization and sequentially numbered, opaque, sealed envelopes. Inadequate methods of allocation concealment included alternation, case record number, dates of birth, day of the week and open lists of random number assignments. When trialists did not report any method of concealment their studies were classified in the review as unclear.

In addition to the allocation concealment score, data were extracted from the review for each trial on the actual method of allocation concealment, the method of generating the sequence for randomization (e.g. random numbers generated by computer, table of random numbers, drawing of lots or envelopes, tossing a coin, shuffling cards or throwing dice) and the method of blinding (Schulz and Grimes, 2002c). If information was available on who was blinded then this was also extracted.

Data analysis

Each meta-analysis included in the sample was then recalculated in Stata using the *metan* command (version 8 for Windows, Stata Corporation, College Station, Texas) and the odds ratio and fixed effect model to produce a pooled effect estimate for each meta-analysis. The *metan* command was used as it allowed the analysis of a trial based on the number of individuals who did and did not experience disease events in the intervention and control groups, as was presented in the original meta-analyses. After completing the recalculation, each meta-analysis was then repeated including only published trials, repeated including only grey trials and finally repeated including only those trials which included published trials supplemented with personal communication. An earlier study by Egger and colleagues (2003) compared the results of the replicated meta-analyses with the original meta-analyses and found them to be the same; this was therefore not done in this study.

To obtain consistency across the meta-analyses the pooled effect estimates were recalculated, where necessary, so that all results were expressed as undesirable outcomes (e.g. mortality not survival, presence of symptoms not absence of symptoms). Thus an odds ratio less than 1.0 indicated that treatments provided a beneficial effect whereas a value greater than 1.0 indicated treatments with an adverse effect.

For each meta-analysis a ratio of the pooled effect estimate from grey trials over the pooled effect estimate of the published trials was calculated. This analysis was also carried out in Stata but this time using the *meta* command which analyses the effect estimates (odds ratio) and corresponding standard errors for each of the meta-

analyses. A ratio greater than 1.0 would indicate that published trials showed a greater treatment effect, likewise a ratio of below 1.0 would indicate that grey trials showed a greater treatment effect. A weighted average was then calculated to combine the ratios across all of these meta-analyses using a random effects model. The random effects model was used to compensate for the perceived heterogeneity between individual meta-analyses. This analysis was then repeated to produce a ratio of the pooled effect estimate of published trials supplemented with personal communication over the pooled effect estimate of published trials. Finally, a ratio of the pooled effect estimate of grey trials over the pooled effect estimate of published trials supplemented with personal communication was calculated. The percentage weight contributed by grey trials, published trials and published trials supplemented with personal communication, in individual meta-analyses were also calculated.

The analysis described above was repeated for the assessment of the effect of methodological quality on the overall results of the meta-analyses. For each meta-analysis a ratio of the pooled effect estimate from trials reporting adequate concealment of the allocation sequence over the pooled effect estimate of trials reporting unclear or inadequate concealment of the allocation sequence was calculated. A weighted average was then calculated to combine these ratios using a random effects meta-analysis. Similarly, a ratio of the pooled effect estimate of trials reporting adequate generation of the allocation sequence over the pooled effect estimate of trials reporting unclear or inadequate generation of the allocation sequence was also calculated. Finally, a ratio of the pooled effect estimate of double blind trials over the pooled effect estimate of other trials was calculated. A ratio

greater than 1.0 would indicate that higher quality trials showed a greater treatment effect likewise a ratio of below 1.0 would indicate that poorer quality trials would show a greater treatment effect.

Results

Description of all 130 systematic reviews

One hundred and thirty systematic reviews of cancer were identified from the 1745 reviews published in the *Cochrane Database of Systematic Reviews* (Issue 3, 2003).

Country of contact author

The contact authors for the 130 reviews (usually the first author of the review) came from 18 different countries and just over half (52%) of these were from the UK.

Table 3:1: Country of contact author of the review

Country of contact author	No. Reviews (%) n=130
Australia	6 (5%)
Brazil	2 (2%)
Canada	8 (6%)
China	1 (1%)
Cuba	1 (1%)
Denmark	8 (6%)
France	1 (1%)
Germany	4 (3%)
Israel	1 (1%)
Italy	3 (2%)
Korea, South	1 (1%)
Netherlands	3 (2%)
New Zealand	3 (2%)
Norway	1 (1%)
Singapore	2 (2%)
Spain	5 (4%)
UK	68 (52%)
USA	12 (9%)

Other countries included the USA (9%), Canada (6%), Denmark (6%), Australia (5%), Spain (4%) and Germany (3%) (Table 3.1). The contact authors for 126 of the 130 reviews (97%) were from high-income countries, two were from a middle- to high-income country (Brazil) and two were from low- to middle-income countries (China and Cuba). This is based on the data and countries classification of the World Bank (www.worldbank.org/data/countryclass/classgroups.htm).

The 130 systematic reviews were from a number of different Cochrane Collaborative Review Groups. The groups containing the greatest number of reviews were the Tobacco Addiction Group (23%), the Gynaecological Cancer Group (21%), the Breast Cancer Group (12%), the Lung Cancer Group (10%) and the Pain, Palliative and Supportive Care Group (10%) (Figure 3.1).

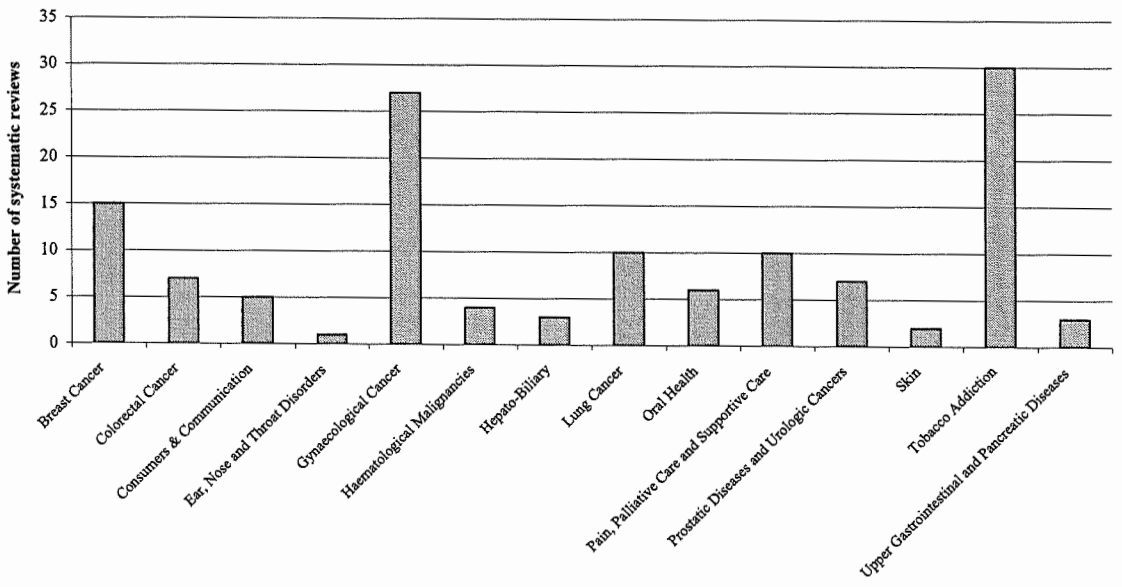


Figure 3:1: Collaborative Review Group and number of systematic reviews

Included studies mentioned in the reviews

The "Included Studies" section of a Cochrane review contains those trials or studies, which met the inclusion criteria and were included in the systematic review. Of the 130 reviews initially assessed, 111 reviews contained published trials in the "Included Studies" section of the review, 40 reviews contained grey trials and 23 reviews contained published trials supplemented with personal communication. Twelve reviews only included individual patient data trials and five reviews did not contain any trials in the "Included Studies" section of the review.

Overall, the median number of trials included in a review was 11 (IQR 6-23), ranging from one to 110 trials per review. The majority of the trials included in the reviews were from published sources; the median number of published trials per review was nine (IQR 5-18). The proportion of grey trials and published trials supplemented with personal communication per review were much smaller. The median number of grey trials per review was two (IQR 1-3) and the median number of published trials supplemented by personal communication was three (IQR 2-7) (Table 3.2).

Table 3:2: Details of included studies (n=130)

Included studies	**Overall**	**Published trials**	**Grey trials**	**Published with personal com.**	**IPD reviews**
Number of reviews	125*	111	40	23	12
Number of trials	2,171	1,628	99	102	342
Median number per review (IQR)	11 (6-23)	9 (5-18)	2 (1-3)	3 (2-7)	13 (7-51)
Range per review (min - max)	1-110	1-107	1-11	1-20	5-66

* Five reviews had no included studies

Excluded studies and studies awaiting assessment mentioned in the reviews

The "Excluded Studies" section of a Cochrane review contains those trials or studies which were considered for the review but do not meet the inclusion criteria. One hundred and twelve reviews included details of excluded studies; the remaining 18 reviews did not. A description of excluded studies was conducted in order to explore whether grey trials were over represented in this section of a Cochrane review. The excluded trials in 109 reviews contained trials from published sources and the median number of published "Excluded" trials per review was 11 (IQR 4-25). Fifty-five reviews listed excluded trials from grey literature sources and the median number of "Excluded" grey trials per review was two (IQR 1-5). There were no published trials supplemented by personal communication in the "Excluded Studies" section of any review (Table 3.3).

Table 3:3: Details of excluded studies and studies awaiting assessment mentioned in the reviews (n=130)

Excluded studies	**Overall**	**Published trials**	**Grey trials**
Number of reviews	112*	109	55
Number of trials	2,338	2,111	227
Median number per review (IQR)	11 (5-28)	11 (4-25)	2 (1-5)
Range per review (min - max)	1-150	1-115	1-50
Studies awaiting assessment	**Overall**	**Published trials**	**Grey trials**
Number of reviews	23**	18	9
Number of studies	116	103	13
Median number per review (IQR)	2 (1-4)	2 (1-4)	1 (1-2)
Range per review (min - max)	1-62	1-62	1-2

*18 reviews had no excluded studies
**107 reviews had no studies awaiting assessment

The "Studies Awaiting Assessment" section of a Cochrane review contains potentially relevant trials or studies that have been identified but cannot be assessed

for inclusion until additional data or information are obtained. Again, it was of interest to explore whether grey trials were over represented in this group, for example due to a lack of information. Of the 130 reviews, 23 contained references to studies awaiting assessment, of these 18 reviews contained references to published trials and nine reviews contained references to grey trials. The median number of trials awaiting assessment was two (IQR 1-4) for published trials and one (IQR 1-2) for grey trials (Table 3.3).

Description of all 25 systematic reviews included in the main analysis

Of the 130 reviews identified in the original sample, 52 reviews were excluded from the main analysis because they did not contain meta-analyses, 12 were excluded because they contained individual patient data meta-analyses, 38 were excluded because only published trials were included in the meta-analyses and three were excluded because only published trials supplemented with personal communication were included in the meta-analyses.

The remaining 25 systematic reviews contained meta-analyses including published trials and grey trials or published trials supplemented by personal communication and were eligible for inclusion in the main analysis (Figure 3.2). Two of the 25 systematic reviews were reviews of complementary medicine (one of acupuncture for smoking cessation and one of self-help interventions for smoking cessation), one was a review of mammography screening for breast cancer and the remaining 22 reviews were all of drug interventions (Table 3.4).

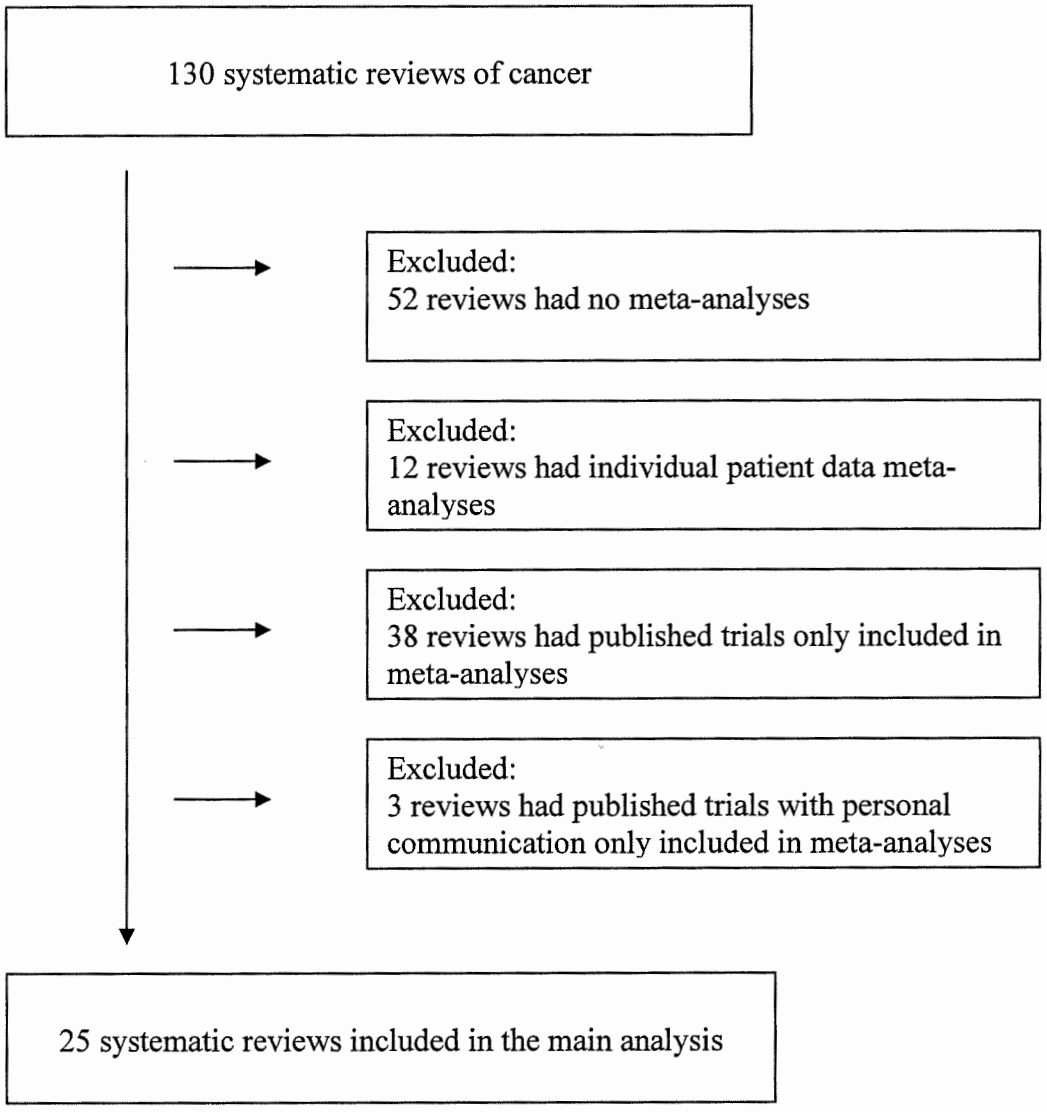


Figure 3:2: Process for identifying eligible systematic reviews for inclusion in the main analysis

Table 3.4: Review titles and the number of trials per meta-analysis

Review	Review title	Overall trials	Published trials	Grey trials	Published with personal com.
Review 1	Bisphosphonates for breast cancer	10	6	1	3
Review 2	Chemotherapy alone versus endocrine therapy alone for metastatic breast cancer	9	8	1	0
Review 3	High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy in women with early poor prognosis breast cancer	8	1	5	2
Review 4	High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy in women with metastatic breast cancer	4	0	3	1
Review 5	Screening for breast cancer with mammography	8	6	0	2
Review 6	Taxane containing regimens for metastatic breast cancer	15	13	2	0
Review 7	Amphotericin B lipid soluble formulations versus amphotericin B in cancer patients with neutropenia	7	5	0	2
Review 8	Amphotericin B versus fluconazole for controlling fungal infections in neutropenic cancer patients	14	8	0	6
Review 9	Beta-lactam versus beta-lactam-aminoglycoside combination therapy in cancer patients with neutropenia	26	17	4	5
Review 10	Concomitant chemotherapy and radiation therapy for cancer of the uterine cervix	11	9	2	0
Review 11	Nystatin prophylaxis and treatment in severely immunodepressed patients	8	6	2	0
Review 12	Routine versus selective antifungal administration for control of fungal infections with cancer	22	13	2	7
Review 13	Short versus long duration infusions of paclitaxel for any adenocarcinoma	2	1	1	0
Review 14	Bisphosphonates in multiple myeloma	10	9	1	0
Review 15	Granulopoiesis-stimulating factors to prevent adverse effects in the treatment of malignant lymphoma	11	1	1	9
Review 16	Interventions for preventing oral candidiasis for patients with cancer receiving treatment	9	5	0	4
Review 17	Corticosteroids for the resolution of malignant bowel obstruction in advanced gynaecological cancer	3	2	1	0
Review 18	Immunotherapy for advanced renal cell cancer	10	9	1	0
Review 19	Maximal androgen blockade for advanced prostate cancer	14	13	1	0
Review 20	Acupuncture for smoking cessation	14	11	0	3
Review 21	Antidepressants for smoking cessation	16	10	6	0
Review 22	Clonidine for smoking cessation	6	5	1	0
Review 23	Nicotine replacement therapy for smoking cessation	51	49	2	0
Review 24	Self-help interventions for smoking cessation	10	8	1	1
Review 25	Combined chemotherapy and radiotherapy (without surgery) compared with radiotherapy alone in localized carcinoma of the esophagus	8	7	1	0

Included studies mentioned in the reviews

Of the 25 reviews included in the main analysis, 24 reviews contained published trials in the "Included Studies" section of the review, 20 reviews contained grey trials and 13 reviews contained published trials supplemented with personal communication. Overall, the median number of trials included in a review was 16 (IQR 10-28), ranging from one to 110 trials per review. The majority of the trials included in the reviews were from published reference sources and the median number of published trials per review was 11 (IQR 8-20). The proportion of grey trials and published trials supplemented with personal communication per review were much smaller. The median number of grey trials per review was two (IQR 1-4) and the median number of published trials supplemented by personal communication was three (IQR 2-8) (Table 3.5).

Table 3:5: Details of included studies (n=25)

Included studies	**Overall**	**Published trials**	**Grey trials**	**Published with personal com.**
Number of reviews	25	24	20	13
Number of studies	569	442	62	65
Median number per review (IQR)	16 (10-28)	11 (8-20)	2 (1-4)	3 (2-8)
Range per review (min - max)	3-110	1-107	1-11	1-11

Excluded studies and studies awaiting assessment mentioned in the reviews

Twenty-four reviews contained trials from published reference sources in the "Excluded Studies" section of the review and 17 reviews contained trials from grey reference sources in this section. Once again there were more published "Excluded" trials per review than grey "Excluded" trials. The median of excluded trials per review was seven (IQR 2-19) for published trials compared to three (IQR 1-4) for grey trials (Table 3.6).

Overall, eight of the 25 reviews contained references to "Studies Awaiting Assessment". Of these, five reviews included references to published trials and six reviews included references to grey trials. Seventeen reviews did not contain any studies awaiting assessment. The median number of trials awaiting assessment per review was two (IQR 1-3) for published trials and one (IQR 1-2) for grey trials (Table 3.6).

Table 3:6: Details of excluded studies and studies awaiting assessment mentioned in the reviews (n=25)

Excluded studies	**Overall**	**Published trials**	**Grey trials**
Number of reviews	25	24	17
Number of studies	325	274	51
Median number per review (IQR)	9 (3-24)	7 (2-19)	3 (1-4)
Range per review (min - max)	1-37	1-37	1-10
Studies awaiting assessment	**Overall**	**Published trials**	**Grey trials**
Number of reviews	8*	5	6
Number of studies	19	10	9
Median number per review (IQR)	2 (1-3)	2 (1-3)	1 (1-2)
Range per review (min - max)	1-5	1-4	1-2

* 17 reviews had no studies awaiting assessment

Trials included in the meta-analyses

The meta-analysis displaying the primary outcome for each of the 25 systematic reviews was assessed (Table 3.4). Overall, the 25 meta-analyses included 311 trials with a median number of 10 trials per meta-analysis (IQR 8-14). This ranged from two to 51 trials per meta-analysis. Twenty-four meta-analyses included published trials (median of eight trials per meta-analysis (IQR 5-11)), 20 meta-analyses included grey trials (median of one trial per meta-analysis (IQR 1-2)), and 12 meta-

analyses included published trials supplemented with personal communication (median of three trials per meta-analysis (IQR 2-6)) (Table 3.7).

Table 3:7: Trials included in the meta-analyses

Trials included in meta-analyses	Overall	Published trials	Grey trials	Published with personal com.
Number of meta-analyses	25	24	20	12
Number of trials	311	224	40	47
Median per meta-analysis (IQR)	10 (8-14)	8 (5-11)	1 (1-2)	3 (2-6)
Range per meta-analysis (min - max)	2-51	1-49	1-6	1-9

Statistical model and type of effect estimate used in the meta-analyses

There was some variation in the type of outcome measures (effect estimate) and statistical methods used to combine the results of the individual trials in the meta-analyses. Of the 25 meta-analyses, 11 used relative risk, nine used the Peto odds ratio and five used odds ratios. The majority of the meta-analyses were performed using a fixed effect model; only five used a random effects model (Table 3.8). This may reflect the default method in the software used by Cochrane reviewers, which is the fixed effect model.

Table 3:8: Statistical model and type of effect estimate used in the meta-analyses

Statistical model	Effect estimate			
	Odds ratio	Peto odds ratio	Relative risk	Overall
Fixed effect	3	9	8	20
Random effects	2	-	3	5
Overall	5	9	11	25

Participants included in the meta-analyses

Overall, the 311 trials included in the 25 meta-analyses contained a median of 155 participants per trial (IQR 76-326). Surprisingly, the grey trials contained a larger

number of participants than the published trials. The median number of participants per trial was 167 (IQR 77-326) for the published trials, compared to 186 (IQR 80-349) for the grey trials; however, this difference was not statistically significant (p=0.6627) using the Mann-Whitley U test. The median number of participants was smallest for published trials supplemented with personal communication (median 100; IQR 56-151) (Table 3.9).

Table 3:9: Participants included in the meta-analyses

Participants included in trials	Overall	Published trials	Grey trials	Published with personal com.
Total number of trials	311	224	40	47
Total number of participants	452,448	346,382	10,162	95,904
Median number per trial (IQR)	155 (76-326)	167 (77-326)	186 (80-349)	100 (56-151)
Range per trial (min - max)	13-76970	13-76970	25-885	25-50430

Sources of the trials included in the meta-analyses

The majority of the 311 trials were published as full journal articles (n=224; (72%)), 47 (15%) were published trials supplemented by personal communication, 26 (8%) were abstracts, 10 (3%) were abstracts supplemented by personal communication, three (1%) were personal communication only and one was a letter (Figure 3.3).

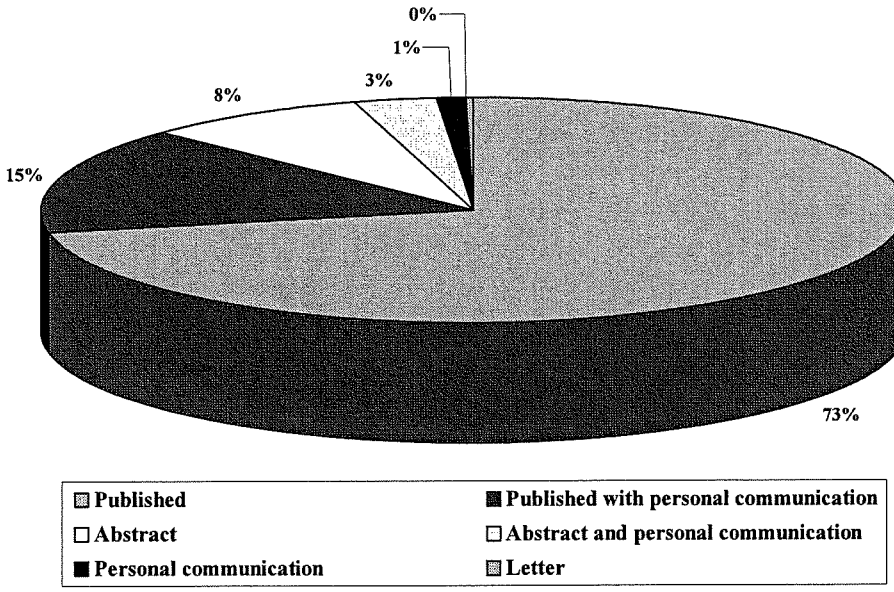


Figure 3:3: Source of trials included in the meta-analyses

Methodological quality of the trials included in the meta-analyses

One aspect of the methodological quality of a trial is the method of allocation concealment. In the software used to conduct Cochrane reviews, the authors of the review are able to score a trial as having either adequate, inadequate or unclear allocation concealment. Overall, 106 (34%) trials were scored as having adequate allocation concealment, eight (3%) trials were scored as having inadequate allocation concealment and for 166 (53%) trials the method of allocation concealment was scored as unclear. The method of allocation concealment was not available for the remaining 31 (10%) trials, as it had not been assessed by the authors of the review (Table 3.10).

There was no significant difference in the number of published trials reporting adequate methods of allocation concealment compared to grey trials (27% versus

30%). However, the number of trials reporting adequate methods of allocation concealment was significantly greater for those trials which were published and supplemented with personal communication (72%). This may well be because the authors of Cochrane reviews sought information on allocation concealment from the authors of the published reports.

Table 3:10: Allocation concealment score

Allocation concealment score	Overall	Published trials	Grey trials	Published with personal com.
Adequate	106 (34%)	60 (27%)	12 (30%)	34 (72%)
Unclear	166 (53%)	131 (58%)	25 (63%)	10 (21%)
Inadequate	8 (3%)	7 (3%)	-	1 (2%)
Not used by reviewer	31 (10%)	26 (12%)	3 (7%)	2 (5%)
Overall	311	224	40	47

Information was also extracted from the review on the actual method of generating the allocation sequence and the concealment of the allocation sequence in each trial. This information was based on the information reported by the authors of the Cochrane reviews for each of the included trials.

Information on the generation of the allocation sequence was available for 221 of the 311 trials. No information was available for 90 trials, as it had not been reported by the authors of the Cochrane reviews. Overall 59 (27%) trials reported adequate generation of the allocation sequence, 11 (5%) trials reported inadequate generation of the allocation sequence and for 151 (68%) trials the generation of the allocation sequence was unclear (Table 3.11). Examples of adequate methods of generation of the allocation sequence included computer randomization, random number tables and the drawing of lots or envelopes. Inadequate methods of generation of the

allocation sequence included allocation by date of birth, clinic group, or by day of the week.

Two hundred and forty-five trials provided information on the concealment of the allocation sequence. No information was available for 66 trials. Eighty (33%) trials reported adequate methods of concealment of the allocation sequence (for example central randomization, coded drug containers prepared in the pharmacy and sealed opaque envelopes). For the remaining 165 (67%) trials, the concealment of the allocation sequence was reported as unclear (Table 3.11).

Overall, there was no difference between the published trials and grey trials in the reporting of the generation of the allocation sequence and concealment of the allocation sequence in the review. However, significantly more published trials supplemented with personal communication reported adequate methods of generation and concealment of allocation compared to the published and grey trials. This may again reflect the type of supplementary information obtained by the authors of the Cochrane reviews from the authors of the published trial reports.

Information on the method of blinding was available for 216 of the 311 trials. The authors of the Cochrane reviews did not provide information on the method of blinding for 79 of the 311 trials and for 16 trials the authors of the review inferred blinding but they did not provide any details. Of the 216 trials in which the authors of the Cochrane review reported on the method of blinding, 53 (25%) trials were reported as double blinded, 11 (5%) trials were reported as single blinded and 63 (29%) trials were open label or not blinded (Table 3.11). For 51 (24%) trials it was

not possible to blind the patient or the clinician (for example these included trials comparing the effects of surgery versus radiotherapy or trials comparing chemotherapy versus radiotherapy) and for the remaining 38 (17%) trials the method of blinding was not reported. Once again, there was no significant difference between the numbers of published trials that were reported as double blinded compared to grey trials (27% versus 29%). Information on who was blinded was available for only ten of the 311 trials.

Table 3:11: Methodological quality of the trials included in the meta-analyses

	Overall	**Published trials**	**Grey trials**	**Published with personal com.**
Generation of allocation sequence				
Adequate	59 (27%)	38 (24%)	7 (21%)	14 (54%)
Unclear	151 (68%)	116 (71%)	26 (79%)	9 (35%)
Inadequate	11 (5%)	8 (5%)	-	3 (11%)
(information not available)	(90)	(62)	(7)	(21)
Concealment of allocation sequence				
Adequate	80 (33%)	46 (26%)	10 (29%)	24 (69%)
Unclear	165 (67%)	129 (74%)	25 (71%)	11 (31%)
Inadequate	-	-	-	-
(information not available)	(66)	(49)	(5)	(12)
Method of blinding				
Double blinded	53 (25%)	40 (27%)	10 (29%)	3 (9%)
Single blinded	11 (5%)	8 (5%)	-	3 (9%)
Unclear	38 (17%)	29 (19%)	4 (12%)	5 (15%)
Not blinded	63 (29%)	40 (27%)	5 (15%)	18 (55%)
Could not be blinded	51 (24%)	32 (22%)	15 (44%)	4 (12%)
(information not available)	(95)	(75)	(6)	(14)

Estimates of treatment effects for published and grey trials

For each of the 25 meta-analyses the odds ratio, as a measure of the precision of the treatment effect, was recalculated in STATA using the fixed effect model (this was the model primarily used in the original Cochrane reviews (Table 3.8)) to produce a pooled effect estimate for each meta-analysis. The odds ratio was then recalculated for each meta-analysis including only published trials, grey trials and then for

published trials supplemented with personal communication (Appendix 2). The odds ratios for published trials could then be compared to the odds ratios for all trials (overall) and changes in statistical significance in each of the meta-analyses could be assessed based on changes in the 95% Confidence Intervals and p-values (Table 3.12).

Table 3:12: Odds ratios for published trials versus overall trials

Review	Published odds ratio (95%CI)	Overall odds ratio (95% CI)	Change in statistical significance
Review 1	1.16 (0.92-1.38)	0.95 (0.72-1.18)	No
Review 2	0.71 (0.41-1.01)	0.71 (0.41-1.01)	No
Review 3	69.84 (4.26-1144.80)	7.92 (5.12-10.72)	No
Review 4	-	5.85 (1.31-26.99)	-
Review 5	0.99 (0.95-1.02)	0.99 (0.96-1.02)	No
Review 6	0.73 (0.59-0.87)	0.74 (0.60-0.88)	No
Review 7	0.77 (0.41-1.13)	0.77 (0.41-1.13)	No
Review 8	0.90 (0.62-1.18)	0.88 (0.59-1.16)	No
Review 9	0.83 (0.56-1.09)	0.83 (0.56-1.09)	No
Review 10	0.57 (0.40-0.75)	0.63 (0.46-0.81)	No
Review 11	0.21 (0.15-0.56)	0.21 (0.14-0.56)	No
Review 12	0.39 (0.20-0.58)	0.40 (0.21-0.59)	No
Review 13	0.06 (0.00-0.56)	0.09 (0.00-0.59)	No
Review 14	0.97 (0.77-1.17)	0.98 (0.79-1.18)	No
Review 15	0.87 (0.25-1.49)	0.87 (0.25-1.49)	No
Review 16	0.50 (0.09-1.10)	0.49 (0.10-1.09)	No
Review 17	0.78 (0.00-2.09)	0.51 (0.00-1.81)	No
Review 18	0.19 (0.00-0.83)	0.16 (0.00-0.79)	No
Review 19	0.87 (0.76-0.99)	0.88 (0.76-0.99)	No
Review 20	**0.73 (0.49-0.96)**	**0.83 (0.59-1.06)**	**Yes**
Review 21	0.52 (0.28-0.75)	0.51 (0.27-0.74)	No
Review 22	0.49 (0.08-0.91)	0.52 (0.11-0.94)	No
Review 23	0.60 (0.51-0.69)	0.60 (0.51-0.69)	No
Review 24	0.76 (0.56-0.96)	0.73 (0.53-0.93)	No
Review 25	0.79 (0.45-1.13)	0.69 (0.35-1.03)	No

There was a change in the statistical significance of the results for published trials compared to those for all trials (overall) in only one of the 25 meta-analyses. This was a meta-analysis of acupuncture for smoking cessation (review 20). In this meta-analysis, published trials showed a statistically significant treatment effect whereas when published trials supplemented with personal communication were included in

the meta-analysis the treatment effect was not statistically significant: OR 0.73 (95% CI 0.49-0.96) for published trials compared to 0.83 (95% CI 0.59-1.06) for all trials. It should be noted that using changes in p-values, as a measure of significance, is only a crude measure of testing the impact of published trials in this type of sensitivity analysis, as it does not assess whether this difference is of clinical importance.

Ratio of estimates of treatment effect for published and grey trials

The ratio of pooled treatment effects from grey trials to those from published trials was calculated for each of the 25 meta-analyses using a random effects model. A ratio greater than one indicated that published trials were more beneficial and likewise a ratio below one indicated that grey trials were more beneficial. The results of this pooled analysis are shown in Figure 3.4. The ratio of treatment effect estimates and 95% Confidence Intervals were plotted and the size of the square reflects the statistical weight contributed by each meta-analysis in the overall pooled analysis. The diamond at the bottom of the forest plot represents the pooled analysis. Eighteen meta-analyses included both grey and published trials with a total of 213 trials, divided into 178 published trials and 35 grey trials. One meta-analysis (review 3) was excluded from the pooled analysis because of its extreme treatment effect. There was no significant difference in treatment effect between grey and published trials. The ratio for grey trials to those of published trials was 1.055 (95% CI 0.836-1.330) showing that pooled estimates of treatment effect from grey trials were 5% less beneficial than published trials. This result is not statistically significant.

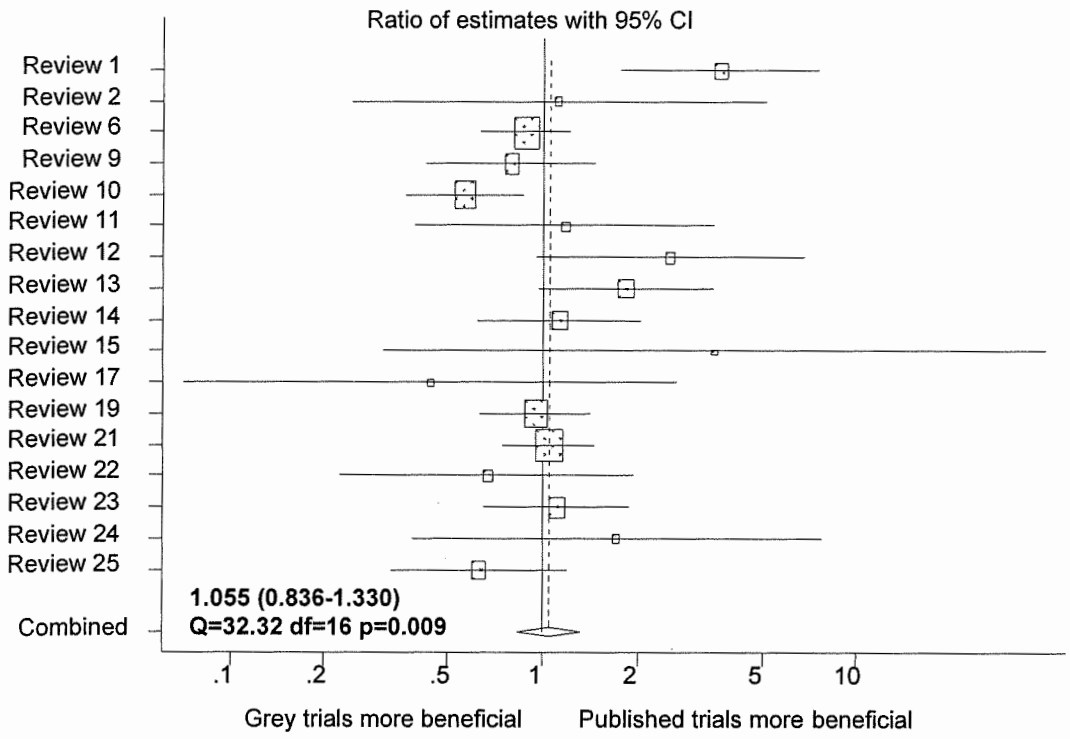


Figure 3:4: Ratio of estimates of treatment effect from grey trials compared to published trials

In this analysis, there was notable heterogeneity between meta-analyses (Q=32, df=16, p=0.009) and wide variation in ratios between different reviews ranging from a ratio of 0.10 (95% CI 0.01-1.93) to a ratio of 3.49 (95% CI 0.30-40.03). Ten meta-analyses showed that published trials were more beneficial, although this difference was only statistically significant for one of these (review 1). This was a review of bisphosphonates for breast cancer, where published trials showed a beneficial treatment effect in improving breast cancer survival whereas the grey trials showed no difference in treatment effect. Seven meta-analyses showed that grey trials were more beneficial; again, this difference was only statistically significant for one of the seven meta-analyses. This difference between meta-analyses does not appear to be explained by the type of intervention under

investigation as all but one were meta-analyses of drug treatments. The only review that did not fall into this category was a meta-analysis of self-help interventions for smoking cessation (review 24).

The analysis was then repeated, excluding those trials reported as conference abstracts from the sample of grey literature trials. The ratio of pooled estimates of treatment effect from grey trials (excluding conference abstracts) to those from published trials was 0.723 (95% CI 0.233-2.249) suggesting that in this analysis grey trials, excluding trials reported in conference abstracts, were actually more beneficial than published trials but, again, this was not statistically significant. However, there was notable heterogeneity between meta-analyses ($p<0.001$) and in this analysis the sample size was extremely small as data were only available from six reviews with a total of 35 trials (25 published trials and 10 grey trials).

Figure 3.5 shows the ratio of the pooled treatment effect estimates for published trials supplemented with personal communication compared to those from published trials. Ten meta-analyses contained relevant data from 120 trials (44 published trials supplemented with personal communication and 76 published trials). One meta-analysis (review 3) was again excluded because of its extreme treatment effect. The ratio for published trials supplemented with personal communication compared to published trials was 1.020 (95% CI 0.911-1.141). This difference between groups was not statistically significant and the results of the pooled analysis were much more homogenous ($Q=8$, $df=9$, $p=0.540$) than before.

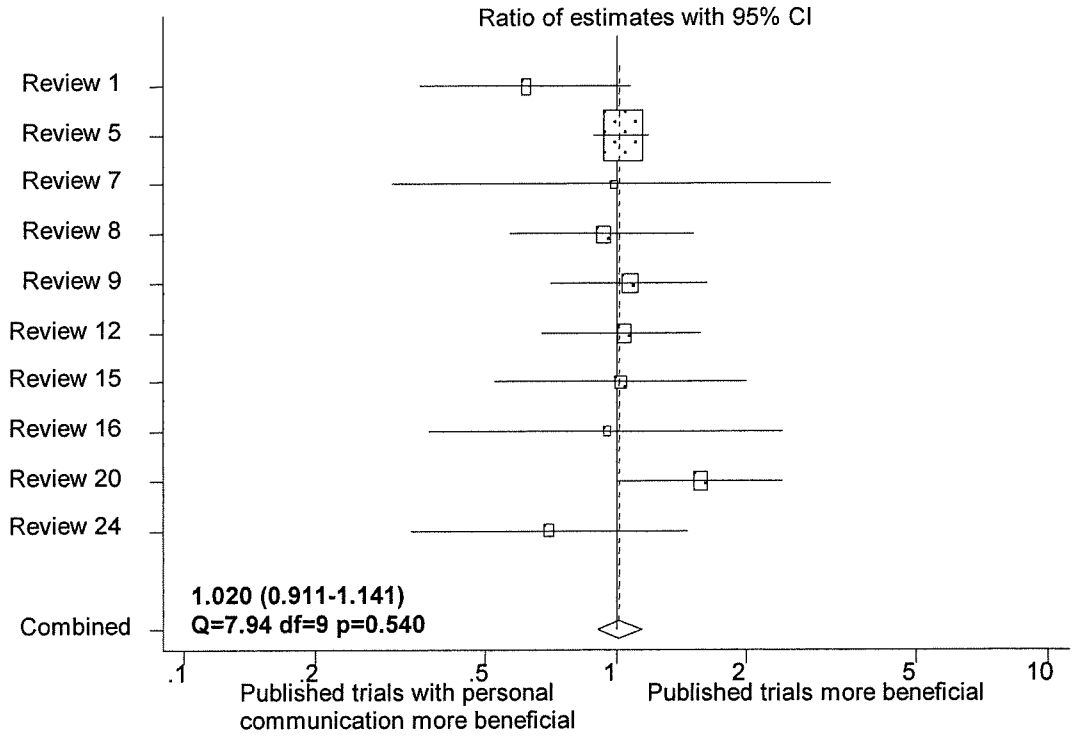


Figure 3:5: Ratio of estimates of treatment effect from published trials supplemented with personal communication compared to published trials

Figure 3.6 shows the ratio of the pooled treatment effect estimates for grey trials compared to those for published trials supplemented with personal communication. Seven meta-analyses contained relevant data from 45 trials (17 grey trials and 28 published trials supplemented with personal communication) and all were included in the analysis. The ratio for grey trials compared to those for published trials was 1.419 (95% CI 0.873-2.306) suggesting that pooled estimates of treatment effect from published trials supplemented with personal communication were 42% more beneficial than those from grey trials. The results of the treatment effect estimates among individual meta-analyses were relatively homogenous (Q=7, df=6, p=0.343). However, this difference between estimates was not statistically significant and represented an analysis of only a small number of trials. Individually, six meta-

analyses showed that published trials supplemented with personal communication were more beneficial and one meta-analysis showed that grey trials were more beneficial, however in none of these individual cases was this difference statistically significant.

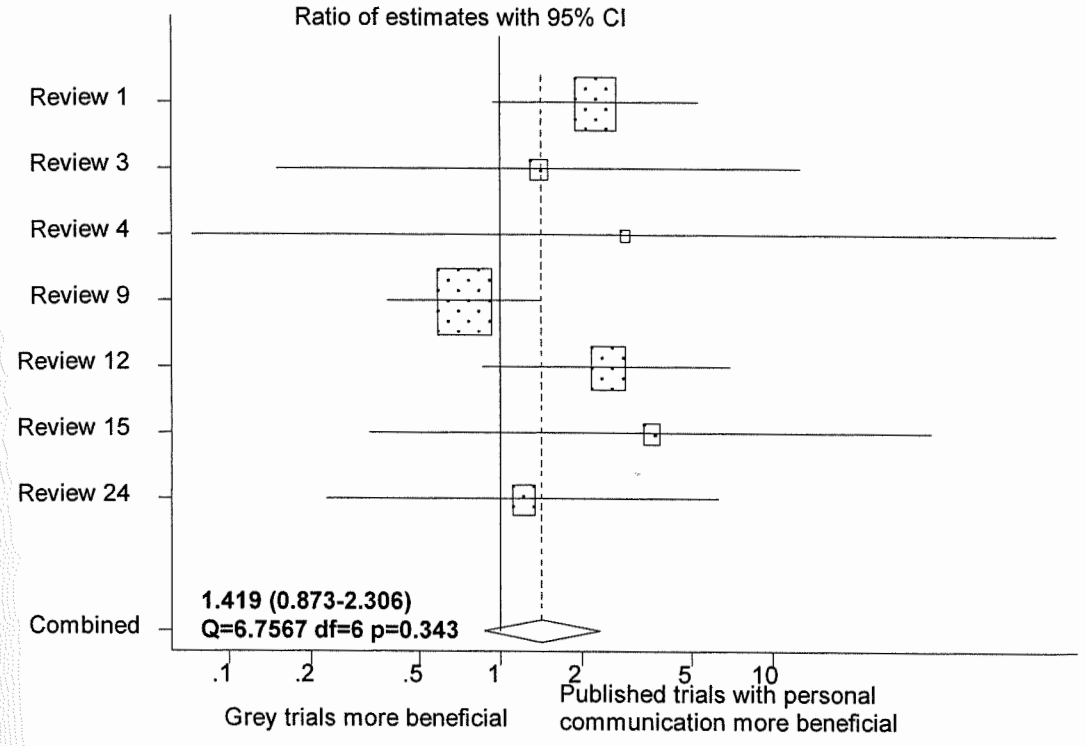


Figure 3:6: Ratio of estimates of treatment effect from grey trials compared to published trials supplemented with personal communication

The size and amount of information contributed by each of the grey trials, published trials and published trials supplemented with personal communication to the results of individual meta-analyses varied greatly. In individual meta-analyses, grey trials contributed a mean of 21% (median 13%) of the overall weight; this ranged from 62% to 2%. Published trials contributed a mean of 74% (median 79%),

ranging from 96% to 15% and published trials supplemented with personal communication contributed a mean of 28% (median 24%), ranging from 83% to 5%.

Estimates of treatment effects for methodological quality

Overall there was no significant difference between the estimates of treatment effect between published trials, grey trials or published trials supplemented with personal communication (Figures 3.4; 3.5; 3.6). The results displayed in Table 3.11 found that there was also no significant difference in the methodological quality of published trials compared to grey trials. Published trials supplemented with personal communication tended to be categorised as being of better methodological quality, which most probably reflects the type of information obtained from the authors of individual trials as this information was often lacking from the reports of published and grey trials.

There is a considerable body of evidence to suggest that trials that are of poorer methodological quality lead to an exaggeration of the effect of treatment (Egger et al, 2003; Kjaergard et al; 2001; Moher et al, 1998; Schulz et al, 1995). A separate analysis of the trials included in this study was therefore carried out to assess the effect of methodological quality on the overall results of the meta-analyses. Figure 3.7 shows the ratio of estimates of treatment effect of trials with adequate concealment of the allocation sequence compared to those with unclear concealment of the allocation sequence. Data were available from 16 reviews with a total of 228 trials. The ratio for adequate trials compared to trials reporting unclear concealment of the allocation sequence was 0.853 (95% CI 0.703-1.034) suggesting

that trials reporting unclear concealment had more beneficial treatment effects, however this difference was not statistically significant. There was also some evidence of heterogeneity between meta-analyses (Q=25, df=15, p=0.045).

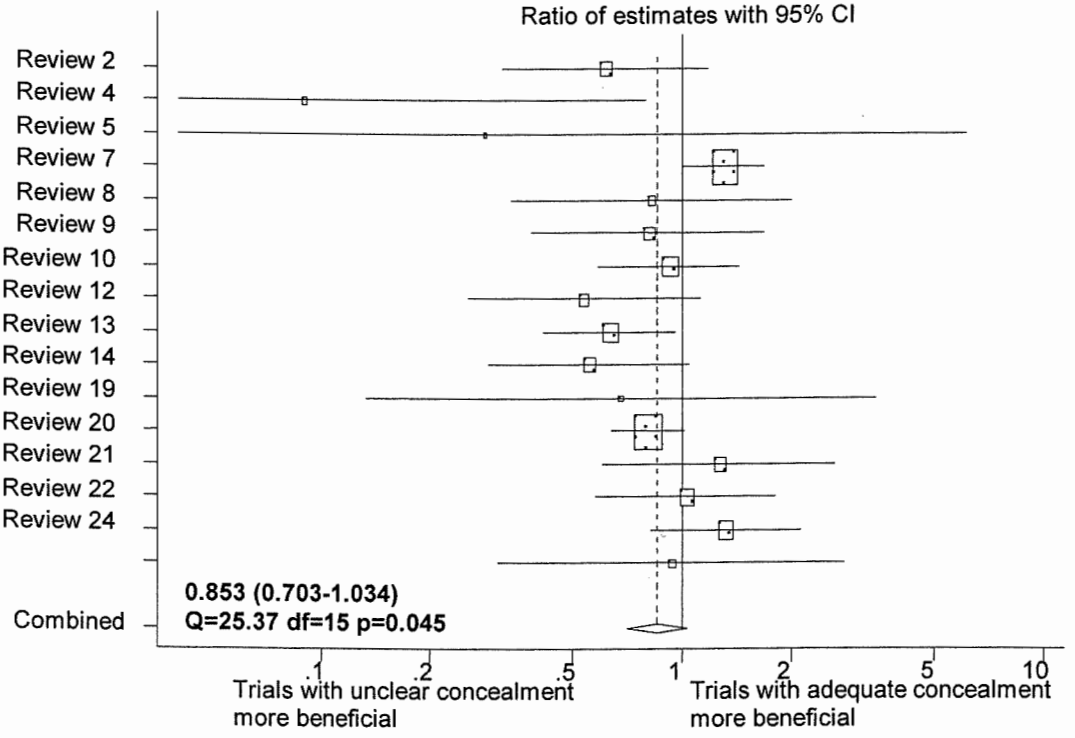


Figure 3:7: Ratio of estimates of treatment effect from trials with adequate versus unclear concealment of allocation sequence

Figure 3.8 shows the ratio of estimates of treatment effect of trials with adequate generation of the allocation sequence compared to those with unclear or inadequate generation of the allocation sequence. Data were available from 16 reviews with a total of 236 trials. Trials reporting inadequate or unclear generation of the allocation sequence were more likely to show beneficial treatment effects than those trials reporting adequate generation of the allocation sequence (ratio of odds ratios 0.872 (95% CI 0.733-1.036)). However, once again, this difference was not statistically

significant and there was some evidence of statistical heterogeneity between meta-analyses (Q=29, df=15, p=0.040).

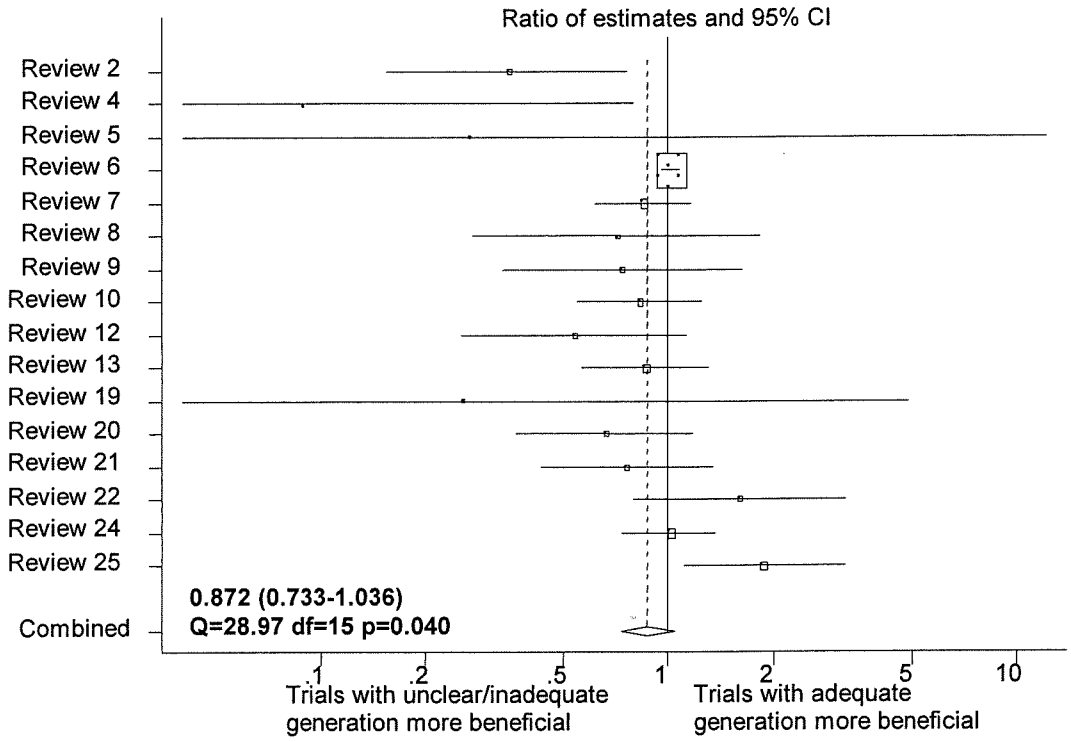


Figure 3:8: Ratio of estimates of treatment effect from trials with adequate versus unclear/inadequate generation of allocation sequence

The ratio of estimates of pooled treatment effects from double blind trials were then compared with those of other trials (Figure 3.9). Trials that were not double blind were more likely to report beneficial treatment effects compared to those which were double blind (ratio of odds ratios 0.755 (95% CI 0.525-1.084)). In this analysis, data were available from seven reviews with a total of 90 trials and there was some evidence of heterogeneity between meta-analyses (Q=14, df=6, p=0.031).

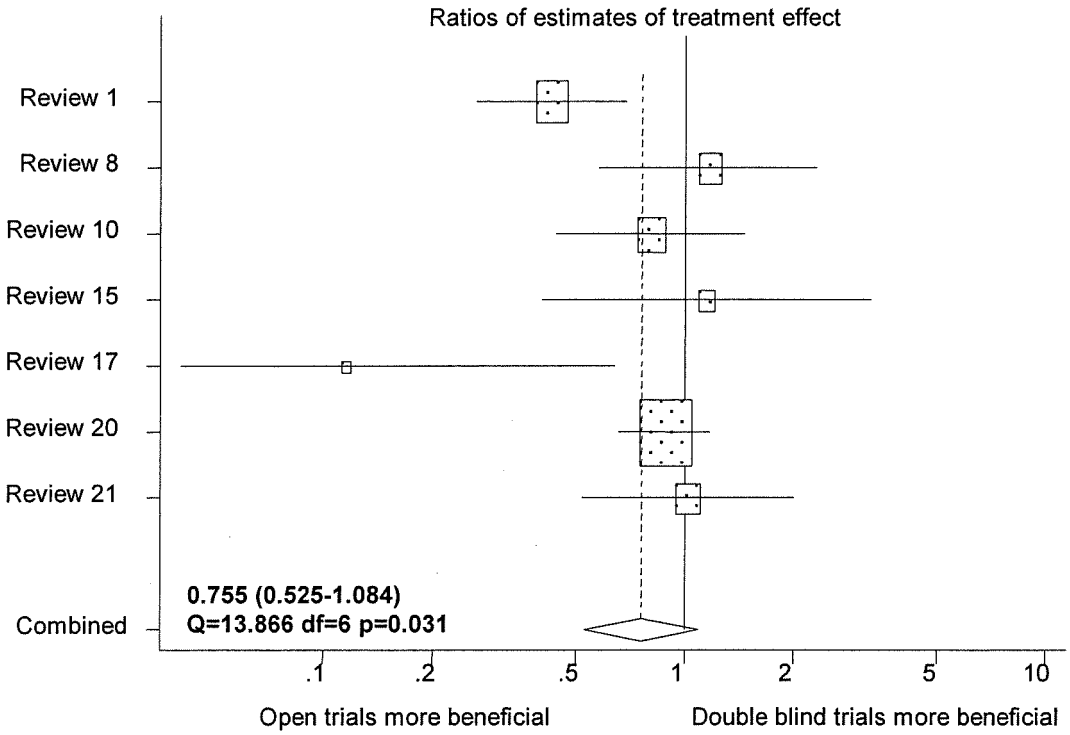


Figure 3:9: Ratio of estimates of treatment effect from double blind trials versus those from other trials

Discussion

Failing to search for, and include, trials found in the grey literature may bias the results and therefore the conclusions of a systematic review. Reassuringly, the results of this study confirm that the authors of Cochrane reviews do appear to search for and include trials from the grey literature in their systematic reviews. As might be expected, of the 130 systematic reviews of cancer published in the *Cochrane Database of Systematic Reviews* included in the original sample, the majority of the reviews included trials referenced by published sources (84%). Encouragingly, just under a third (31%) of the reviews also included trials referenced only by grey literature sources while 18% of reviews included trials

referenced only by published sources that were supplemented with personal communication from the trialists.

One of the potential problems of including trials found in the grey literature in a systematic review is the limited amount of information available about the trial (Hopewell and Clarke, 2003). It is therefore possible that the authors of Cochrane reviews might search for such trials but not be able to include them in the review because of insufficient information and therefore cite them as an excluded study or a study awaiting assessment. This problem was not apparent in this study as the proportion of published and grey trials referenced as excluded studies or studies awaiting assessment were similar to those of trials actually included in the review.

Twenty-five of the 130 systematic reviews, included in the original sample, contained meta-analyses including published trials, grey trials, or published trials supplemented with personal communication. The majority were meta-analyses of drug interventions for the treatment of cancer. Only two were meta-analyses of complementary therapies. It is important to take this into account when considering the findings of this study as there is evidence to suggest that the impact of excluding trials found in the grey literature may be greater in some areas of health care than in others (Egger et al, 2003).

The average number of trials per meta-analysis was ten ranging from two to 51 trials per meta-analysis. This figure is representative of other such studies (described in Chapter Two) where the average number of trials per meta-analysis was 11 in the study by McAuley and colleagues (2000) and 13 in the study by

Egger and colleagues (2003), which specifically excluded meta-analyses containing fewer than five trials.

The majority of the trials included in the meta-analyses were reported by published sources (72%), or published sources supplemented with personal communication (15%). Thirteen per cent of trials were reported only by grey literature sources, of which abstracts were the most common type. Again, these findings are consistent with those in the studies described in Chapter Two, the exception being those published trials where additional information was obtained from the trialists. This clearly reflects The Cochrane Collaboration's policy of encouraging authors of Cochrane reviews to seek more information from trialists where the information required is either not reported or is unclear (Alderson et al, 2003). This distinction is not apparent in other studies that have analysed the use of grey literature in systematic reviews. Interestingly, the number of participants included in the meta-analyses were similar for both published and grey trials. This is different from other such studies (Burdett et al, 2003; Egger et al, 2003; Fergusson et al, 2000; McAuley et al, 2000) where published trials tended to be much larger than grey trials. Reasons for this difference are unclear and may be specific to the sample of trials assessed in this study.

In this present study there was no statistically significant difference between the results of published trials and those of grey trials. The pooled ratio of estimates of treatment effect from published trials were 5% more beneficial (ratio of odds ratios 1.055 (95% CI 0.836-1.330)) than those from grey trials with significant heterogeneity and variation of effect estimates between meta-analyses. This

variation in the pooled ratios of estimates of treatment effect from published and grey trials could not be explained by the type of intervention under investigation, as the majority of the meta-analyses were of drug interventions. The analysis of pooled estimates from published trials compared to those from published trials supplemented with personal communication found no significant difference in treatment effects between these two groups. In comparison the analysis of grey trials compared to published trials supplemented with personal communication showed that published trials supplemented with additional information from the trialists yielded 42% larger estimates of treatment effect, although this was not statistically significant.

This difference in treatment effects between published and grey trials is not as great as those reported by Egger and colleagues (2003) who found that the ratio of estimates of treatment effect from published trials were 7% more beneficial (ratio of odds ratios 1.07 (95% CI 0.98-1.15)) than those from grey trials. Likewise the study by McAuley and colleagues (2000) found that pooled estimates of treatment effect from published trials were 15% more beneficial (ratio of odds ratios 1.15 (95% CI 1.04-1.28)) than grey trials. However, in both of these studies, there was also significant heterogeneity between meta-analyses in different medical areas. In some areas the effect of excluding grey trials from the meta-analysis resulted in greater differences in treatment effect than in other areas. Egger and colleagues (2003) stratified their analysis by area of health care. They identified only four meta-analyses in cancer and found no significant difference in treatment effect between published and grey trials (ratio of odds ratios 0.96 (95% CI 0.86-1.09). It has also been suggested that excluding trials reported as abstracts from the grey

literature could mean that published trials show an even greater treatment effect (McAuley et al, 2000). However, the sample of trials in this present study was too small to provide a meaningful analysis of this point.

In general, the findings of this present study of meta-analyses in cancer support those of the studies identified in Chapter Two of this thesis (Burdett et al, 2003; Egger et al, 2003; Horn and Limburg, 2002; Bhandari et al, 2000; Fergusson et al, 2000; McAuley et al, 2000; Man-Son-Hing et al, 1998; Jeng et al, 1995) which showed there may be differences between the results of published trials compared to those found in the grey literature. Therefore, excluding trials found in the grey literature from a systematic review may lead to an exaggeration of the effect of treatment. However, the findings of this present study also show that published trials do not always yield larger treatment effects than grey trials. Sometimes the effect may be reversed and grey trials may yield larger treatment effects than published trials or, alternatively, there may be no systematic difference between the results of either group of trials. This means that those carrying out systematic reviews should always endeavour to search for trials in both the grey and published literature and, where appropriate, include them in their review. It is not until such trials have been identified and included in a systematic review that their effect on the overall results and, therefore, the conclusion of the systematic review can be analysed.

One of the key aspects of a trial, irrespective of how it is identified, is its methodological quality and how this can affect the overall results of a systematic review. In this study, trials with unclear methods of generation and concealment of

the allocation sequence were more likely to yield larger estimates of treatment effect than those trials with adequate generation and concealment (concealment of allocation sequence: ratio of odds ratios 0.853 (95% CI 0.703-1.034); generation of allocation concealment: ratio of odds ratios 0.872 (95% CI 0.733-1.036)). Likewise, unblinded trials were more likely to yield larger estimates of treatment effect than those trials which were double blinded (ratio of odds ratios 0.755 (95% CI 0.525-1.084)).

These findings support those of previous studies, which show that trials of poorer methodological quality are more likely to yield larger estimates of treatment effect (Jüni et al, 2000; Kjaergard et al, 2001; Moher et al, 1998; Schulz et al, 1995). The findings from these studies and those from this analysis of trials included in Cochrane systematic reviews of cancer have been combined in a meta-analysis (Figure 3.10). In this combined meta-analysis, trials with inadequate or unclear concealment of the allocation sequence (ratio of odds ratios 0.73 (95% CI 0.65-0.83) and lack of double blinding (ratio of odds ratios 0.85 (95% CI 0.75-0.96) were statistically significantly more likely to yield larger estimates of treatment effects than those trials reporting adequate methods. Likewise, trials reporting inadequate or unclear methods of generation of the allocation sequence were more likely to produce larger estimates of treatment effect, although this difference was not statistically significant (ratio of odds ratios 0.85 (95% CI 0.72-1.01). This reinforces the need to routinely assess the quality of individual trials included in a systematic review and their impact on the overall results of a meta-analysis.

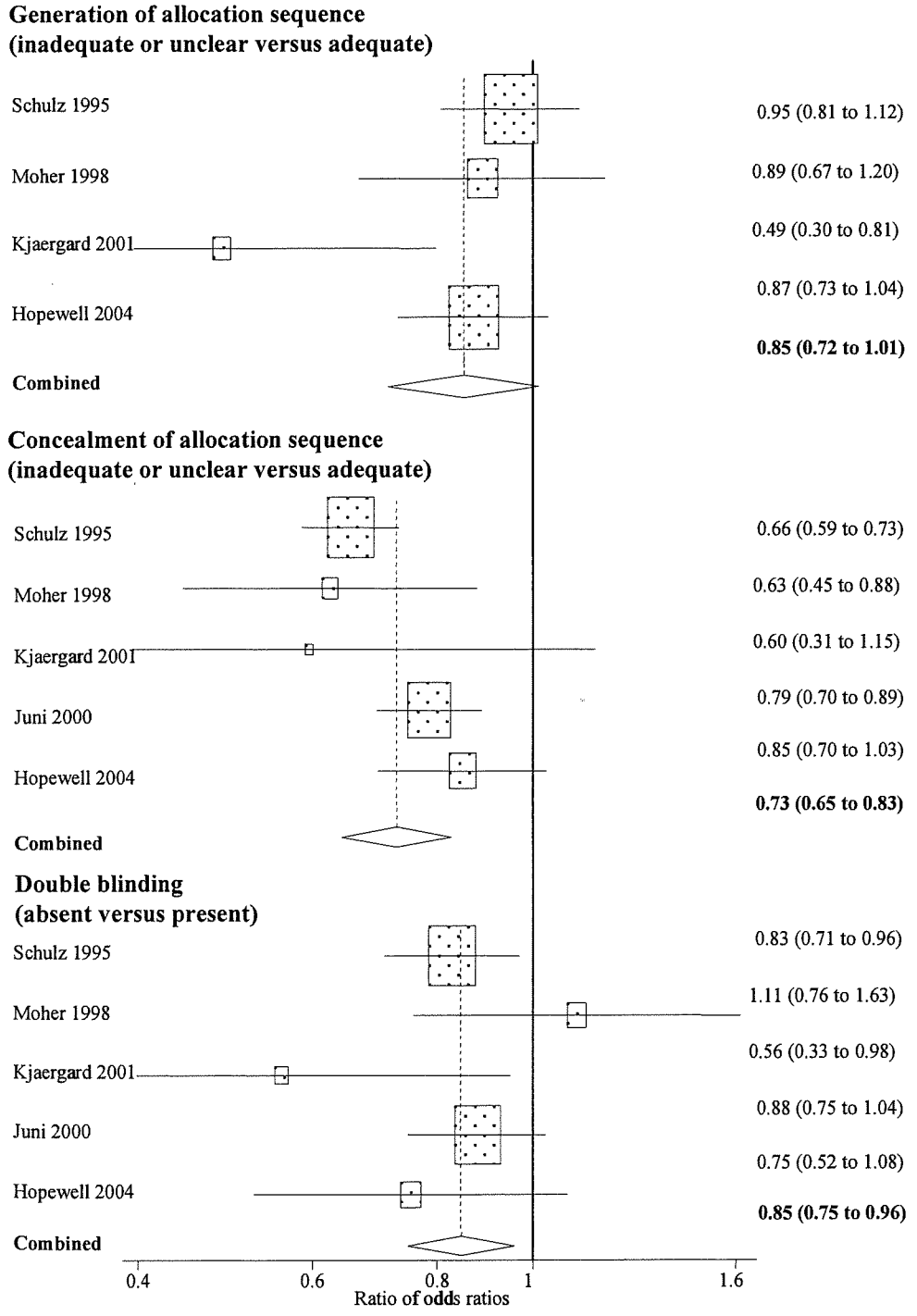


Figure 3:10: Meta-analysis of studies assessing the methodological quality of randomized trials

In the study described in this chapter, methodological quality was assessed in two ways. The first used the allocation concealment score assigned by the authors of the Cochrane review. The majority of published and grey trials were scored as having unclear methods of allocation concealment. In comparison, the majority of published trials supplemented with personal communication were scored as adequate. Egger and colleagues (2003) conducted a similar analysis of trial quality using the allocation concealment score assigned by the authors of the Cochrane reviews, as described in Chapter Two. In that analysis Egger and colleagues (2003) grouped trials with inadequate or unclear allocation concealment together and concluded that grey trials were of poorer methodological quality. However, a recent study by Middleton in 2003 has since shown that this scoring system is used inconsistently by the authors of Cochrane reviews and highlights problems in using it to assess methodological quality. Pildal (2004) has since also highlighted problems with the consistency of the allocation concealment score assigned in Cochrane reviews. In an independent assessment of 37 Cochrane reviews, Pildal found that 95 of the 200 trials assessed were coded incorrectly by the authors of the review when compared with what was reported in the published paper.

The second method of assessing methodological quality, described in this chapter, was to look in more detail at the actual methods of the generation and concealment of the allocation sequence and method of blinding. Again, this method of assessing quality has its limitations as it was dependent on the information reported by the authors of the Cochrane review and was not always assessed. This may point to a need for a more independent assessment of trial quality and publication status. However, it offers an alternative approach to an overall combined score as it allows

more detailed assessment of the individual quality components. In those reviews in which quality had been assessed, the majority of published and grey trials had unclear methods of generation and concealment of the allocation sequence, whereas the majority of published trials supplemented with personal communication had adequate methods. Where the method of blinding had been reported by the authors of the Cochrane reviews, there was no significant difference between grey and published trials, while more detail was provided for published trials supplemented with personal communication. These findings highlight once again the importance of contacting trialists for more information about the methodological quality of the trial. Better awareness of the importance of these issues is needed. This may be especially important for those trials published prior to the introduction of the CONSORT statement in 1996 (Begg et al, 1996), which was revised in 2001 (Moher et al, 2001b) and requires that this information be reported for trials published in journal articles.

Limitations of the existing evidence

Published trials are more likely to show a greater treatment effect than trials found in the grey literature. However, evidence suggests that the extent of this effect varies depending on the type of intervention or healthcare condition under investigation. These findings suggest that those carrying out systematic reviews should always search for trials in both published and grey literature and, where appropriate, include them in the systematic review. However, further evidence is needed to understand better the practical implications, for those carrying out systematic reviews, for finding and trying to include trials published in the grey literature in systematic reviews.

CHAPTER 4: POTENTIAL IMPACT OF INCLUDING GREY LITERATURE IN SYSTEMATIC REVIEWS

Introduction

The evidence presented so far in this thesis shows that trials found in the grey literature can have systematically different results to trials found in the published literature. If only those reports of research that are published in full are included in a systematic review then the results and conclusions of that review could be biased. The inclusion of grey literature (for example conference abstracts, research reports, dissertations, policy documents and personal correspondence) in systematic reviews may help overcome this problem.

The identification of relevant trials in the grey literature can be particularly time-consuming and difficult. The review of the literature presented in Chapter Two showed that approximately half of the trials found in the grey literature were published as abstracts in the proceedings of scientific meetings. A recent study has also shown that approximately 20% of Cochrane reviews contain at least one trial for which the only source was a conference abstract (Mallett et al, 2002). Trials reported in conference abstracts are a primary source of grey literature (McAuley et al, 2000). They are often available only in proceedings or journal supplements and are rarely indexed in databases such as MEDLINE, making them particularly difficult to identify. Even when identified, their inclusion in a systematic review can still be difficult due to the limited amount of information available. Chapter Four of this thesis aims to determine the relative benefits of trying to include a

previously unidentified source of grey literature in systematic reviews and to assess its potential impact on the overall results of a review.

Objectives

To assess the potential impact of including a sample of grey literature in systematic reviews published in *The Cochrane Database of Systematic Reviews* in *The Cochrane Library.*

The first part of this study was a pilot study assessing the potential impact of including a sample of grey literature in systematic reviews. This was then followed by the main study, which aimed to build on the results of the pilot study. As the results of both projects proved to be very similar they are presented together in this chapter. The results of the pilot study have been published in the *Internal Medicine Journal* (Hopewell and McDonald, 2003) and in *Respirology* (Hopewell, 2003) and one aspect of the main study has also been presented at the 11th International Cochrane Colloquium in Barcelona, Spain, in 2003 (Hopewell et al, 2003).

Source of data

For the pilot study the sample of grey literature comprised abstracts presented at the annual conference of the Thoracic Society of Australia and New Zealand between 1981 and 1998 and published in the *Australian and New Zealand Journal of Medicine.* The journal, which was first published in 1952 as the *Australasian Annals of Medicine*, changed to the *Australian and New Zealand Journal of Medicine* in 1971 and began including the proceedings of several medical societies in 1980. In 2001 the journal was renamed the *Internal Medicine Journal.*

For the main study, the sample of grey literature comprised abstracts presented at the annual conference of the American Society of Clinical Oncology in 1992. The American Society of Clinical Oncology is a worldwide organisation representing clinicians and other professionals treating people with cancer. The annual American Society of Clinical Oncology conference is an international meeting, which in 2003 was attended by over 26,000 people from all over the world (www.asco.org). Current research presented at the meeting is compiled and published as abstracts in the proceedings of the conference.

These samples from the Thoracic Society of Australia and New Zealand and the American Society of Clinical Oncology conferences were selected because of the high number of trials reported in the conference proceedings and the expected high number of potentially eligible Cochrane systematic reviews covering these medical areas. These abstracts were also not yet included in the *Cochrane Central Register of Controlled Trials* and therefore were not easily accessible to the authors of Cochrane reviews. This meant that they were likely to be representative of grey literature requiring a specific grey literature search.

Methods

Handsearching

As part of a larger project (outside the scope of this thesis) the abstracts from all conference proceedings published in the *Australian and New Zealand Journal of Medicine* between 1980 and 2000, and all conference abstracts published in the proceedings of the American Society of Clinical Oncology annual conference in

1992 were handsearched to identify all reports of randomized controlled trials using the definition used by The Cochrane Collaboration. This definition requires that "the participants (or other units) were definitely or possibly assigned to one of two (or more) alternative forms of health care using a process of random allocation". Those reports of trials that used a quasi method of randomization (such as alternation, date of birth or case record number), or where randomization could not be ruled out, were classified as controlled clinical trials. All reports judged to be potentially relevant by the handsearcher were then individually assessed, as part of this thesis, to confirm that they met the Cochrane eligibility criteria.

Assessing full publication

For this thesis, each of the trials reported in the conference abstracts were assessed to determine if they had been published as a full article. *The Cochrane Central Register of Controlled Trials* (*CENTRAL*) was used to search for the first author (and then second, third or fourth author if this was not successful) listed in the abstract (Issue 1, 2002 of *The Cochrane* Library was searched in the pilot study and Issue 4, 2002 was searched in the main study). Where a large number of hits were retrieved, a key word from the title was used to help refine the search. The search was then repeated using the U.S. National Library of Medicine's PubMed database (www.ncbi.nlm.nih.gov/entrez/query.fcgi) for those conference abstracts for which a full report was not found in *CENTRAL* (1966 to March 2002 was searched in the pilot study and 1966 to December 2002 was searched in the main study). Possible publications were examined to see if the full report corresponded with the subject matter in the conference abstract. Since *The Cochrane Central Register of Controlled Trials* contains over 150,000 reports of trials from sources other than

MEDLINE (including over 60,000 reports unique to EMBASE), it was decided not to extend the search to other databases beyond MEDLINE. The time interval between the year of the conference, at which the relevant conference abstract had been presented and full publication of the trial's results was also recorded.

Assessment of trial characteristics

Information was extracted from each conference abstract on the number of participants randomized in each trial. The methodological quality of each trial reported in the conference abstract was also assessed based on the method of allocation concealment, which is the process used to prevent foreknowledge of group assignment in the randomized controlled trial. Each trial reported in the conference abstract was assessed as having either adequate allocation concealment (e.g. central randomization office) or inadequate (e.g. date of birth or case record number) or unclear allocation concealment. This method of assessing trial quality was used because it is the method most commonly used by Cochrane reviewers.

Searching for possible systematic reviews

A search was carried out of the *Cochrane Database of Systematic Reviews* (Issue 1, 2002 was searched in the pilot study and Issue 4, 2002 was searched in the main study) for each trial reported in a conference abstract to try to identify systematic reviews for which the conference abstract might be eligible for inclusion. This was done by reading each conference abstract and then using the search function in *The Cochrane Library* to search for keywords relevant to the content of the conference abstract. Retrieved reviews were assessed by reading the title and abstract of the review, to determine whether the trial might be relevant for inclusion.

Where a possibly suitable Cochrane review was identified, its list of references was checked to see if the trial reported in the conference abstract was already included (either as an abstract or as a full article). If the trial was not already included in the review, the inclusion criteria for the review were assessed to see if the trial was eligible. This was done using the 'Criteria for Considering Studies' section of the review, and by assessing the following eligibility criteria:

- types of studies (e.g. the study design).
- types of participants (e.g. the disease or condition of interest and the population and setting).
- types of interventions (e.g. the intervention of interest and what it is being compared against).
- types of outcome measures (e.g. the specific outcomes which are being assessed).

If it was unclear whether the trial reported in the conference abstract met the full eligibility criteria for a review, the contact author of the review was contacted and a copy of the conference abstract was enclosed with a request for their help to determine its eligibility.

Results of the pilot study

This section contains the results of the pilot study assessing the impact of including abstracts from the proceedings of the Thoracic Society of Australia and New Zealand in Cochrane reviews.

Full publication

A total of 962 reports of controlled trials were identified from searching the conference abstracts published in the *Australian and New Zealand Journal of Medicine* between 1980 and 2000. Of these, 726 (76%) were reports of randomized controlled trials and 236 (24%) were reports of quasi-randomized or possibly randomized trials. The median number of trials published per year was 43 (IQR 28-59).

Trials were identified from the conference proceedings of 17 medical societies. Sixty-one percent (n=589) of the trials initially presented as abstracts were subsequently published as full reports (Table 4.1). This varied between medical societies, from 46% for the Australian and New Zealand Society of Nuclear Medicine to 79% for the Australian Society of Neurologists. Thirty-one of the 589 trials, where it was possible to identify full publications, were retrieved from searching PubMed but were not identifiable in the *Cochrane Central Register of Controlled Trials*. As of May 2002, the remaining 373 (39%) abstracts remained unpublished. The average time to publication from presentation at the conference to full publication of results was one to two years, ranging from less than one year to 10 years; 84% were published within two years, 95% published within four years and 99% published within eight years. Time to publication was measured from the year the abstract was presented at the conference to the year of full publication of its results. Four trials had already been published in full one year before being presented as an abstract at a conference.

Table 4:1: Number of trials reported as abstracts in the Australian and New Zealand Journal of Medicine and full publication of their results

Medical Society*	Number of trials	Number published in full (%)
Cardiac Society of Australia and New Zealand	338	221 (65%)
Thoracic Society of Australia and New Zealand	187	101 (54%)
Gastroenterological Society of Australia	138	89 (64%)
Australian Rheumatology Association	100	57 (57%)
Haematology Society of Australia	42	23 (55%)
Australian Society for Geriatric Medicine	38	27 (71%)
Australian and New Zealand Society of Nuclear Medicine	28	13 (46%)
Australian Association of Neurologists	19	15 (79%)
Australasian Society for Infectious Diseases	18	13 (72%)
Australian Society of Nephrology	17	9 (53%)
Australian Diabetes Society	14	8 (57%)
Royal Australasian College of Physicians	11	8 (73%)
Others	12	6 (50%)
Total	**962**	**589 (61%)**

* The name of the medical society may have changed over time.

For the purposes of the pilot study, only the 187 reports of randomized trials presented as abstracts at the annual conference of the Thoracic Society of Australia and New Zealand between 1981 and 1998 were assessed. This sample was selected because of the large number of trials identified at this conference and the perceived high number of potentially eligible Cochrane reviews. Of the 187 reports of randomized trials identified, 101 (54%) had been published as a full article. The average time to full publication of the trials' results was one to two years, ranging from less than one year up to six years. As of May 2002, 86 (46%) reports of randomized trials presented as conference abstracts remained unpublished.

Trial characteristics

Of the 187 reports of randomized trials identified from the proceedings of the Thoracic Society of Australia and New Zealand, 134 (72%) were reports of

randomized controlled trials and 52 (28%) were reports of quasi-randomized or possibly randomized trials. One hundred and eighty one abstracts included information on the number of randomized participants and the total number of participants included in these trials was 9691, ranging between four and 1203 participants per trial. The median number of participants per trial was 20 (IQR 11-47). This was similar for those trials reported in conference abstracts that had subsequently been published as a full report (median 20; IQR 12-48) and was slightly more than that for trials published only as conference abstracts (median 16; IQR 10-46) (Table 4.2). The method of allocation concealment could be determined in only four of the trials reported in the conference abstracts. It was not possible to determine the method of allocation concealment used in the majority of the trials because not enough information was provided in the abstract.

Table 4:2: Characteristics of the trials reported in the Thoracic Society of Australia and New Zealand abstracts

	Overall	**Published in full**	**Published as abstracts**
Total number of abstracts	187	101	86
Year of publication of abstracts	1981-1998	1981-1998	1981-1998
Number of RCTs	134	76	58
Number of CCTs*	53	25	28
Total number of participants**	9691	5903	3788
Median number of participants (IQR)	20 (IQR 11-47)	20 (IQR 12-48)	16 (IQR 10-46)
Range of participants	4-1203	5-1203	4-370
Allocation concealment	Adequate = 4 Unclear = 183	Adequate = 3 Unclear = 98	Adequate = 1 Unclear = 85

* CCTs refers to those reports of trials that used a quasi method of randomization (such as alternation, date of birth or case record number), or where randomization could not be ruled out and were coded as controlled clinical trials

**Data for the number of participants is given for 181 abstracts, six abstracts did not include this information (two published and four unpublished).

Identifying possible systematic reviews

No possibly relevant Cochrane review was found for 145 of the 187 trials reported in the conference abstracts (Table 4.3). Possible reviews (28 reviews) were identified for 42 trials. Twenty-four of these trials were already mentioned in Cochrane reviews (14 reviews): 22 referenced the trial as it had been reported in the full publication (four of which were listed as an "excluded study" by the reviewers) and two referenced the trial as reported in the conference abstract.

Table 4:3: Trials reported in the Thoracic Society of Australia and New Zealand abstracts for which there was a possible Cochrane review

	Overall (n=187)	**Published in full (n=101)**	**Published as abstracts (n=86)**
Possible review identified:	42	28	14
already in review	24	22*	2
said to be eligible by reviewer	3	0	3
said to be ineligible by reviewer	15	6	9
No review identified	145	73	72

*Four abstracts published in full were initially thought to be eligible for inclusion in the review but had subsequently been listed as an excluded study by the authors of the review.

For the remaining 18 trials reported in conference abstracts, it was unclear if they met the full eligibility criteria for the review and the author of each review was contacted for help. If this was not possible, an expert in the field was contacted for advice. Only three of the 18 trials reported in conference abstracts were said to be eligible for inclusion in a review (one author said they would need to contact the authors of the trial for clarification); the remaining 15 trial reports were said not to be eligible for inclusion. The reasons given as to why the trial reports were not eligible for inclusion in the review included: no clinical data available (one); the

types of participants were different to those assessed in the review (one); and the trial compared an intervention which was not assessed in the review (six).

Results of the main study

This section contains the results of the main study, which assessed the impact of abstracts from the proceedings of the American Society of Clinical Oncology conference in 1992.

A total of 209 abstracts containing reports of randomized trials were identified from searching the conference abstracts published in the American Society of Clinical Oncology in 1992. Trials were identified in a number of different cancer specialities (Table 4.4).

Table 4:4: Number of trials reported as abstracts at the American Society of Clinical Oncology conference and full publication of their results

Cancer speciality	Number of trials	Number published in full (%)
Breast cancer	46	27 (59%)
Gastrointestinal cancer	28	17 (61%)
Genitourinary cancer	17	11 (65%)
Gynaecologic cancer	11	4 (36%)
Head and neck cancer	6	5 (83%)
Leukaemia adult	12	7 (58%)
Leukaemia paediatric	5	3 (60%)
Lung cancer	22	12 (55%)
Lymphoma adult	11	11 (100%)
Lymphoma paediatric	4	2 (50%)
Multiple cancers	6	2 (33%)
Non-specific cancer*	30	16 (53%)
Paediatric solid tumours	4	2 (50%)
Sarcoma	3	3 (100%)
Skin cancer	4	3 (75%)
Total	**209**	**125 (60%)**

*Examples of non-specific cancer included trials of pain management, anti-emetics and quality of life.

Abstracts were also identified that mentioned a trial, which might have been randomized, but the results were not the main focus of the abstract. For example, 18 abstracts described a call for a trial to be conducted in a specific area, 18 abstracts described a plan to conduct a trial, 12 abstracts described a trial that had previously been conducted and two abstracts described a meta-analysis of randomized trials. In addition, 22 abstracts, which were thought to contain possible reports of randomized trials were later excluded following closer inspection of the full publication (Hopewell et al, 2003). Of these, 11 described the results of a trial that might have been randomized; five of these had subsequently been published in full and, of these, three were randomized trials. Seven conference abstracts (containing the reports of 14 "possible" randomized trials) mentioned trials, which might have been randomized but the results of these trials were not the main focus of the abstract. Subsequent full publications were found for 11 of these "possible" trials and, of these, eight were randomized trials. Three conference abstracts suggested that a trial, which might be randomized, was being, or should be, undertaken; no subsequent publications were identified for these. One conference abstract included the phrase "randomly selected". This study had subsequently been published in full but it was not a randomized trial.

Full publication

One hundred and twenty five (60%) of the 209 trials initially presented as abstracts were subsequently published as full reports (Table 4.4). Full publications were identified for 116 (55%) of the 209 trials by searching the *Cochrane Central Register of Controlled Trials*. In addition, full publications for nine (4%) of the 209

trials were retrieved by searching PubMed and were not identifiable in the *Cochrane Central Register of Controlled Trials*. As of December 2002, the remaining 84 (40%) abstracts remained unpublished.

The median time to publication from presentation at the conference to full publication of results was 27 months (IQR 15-43 months), ranging from less than one year to eight years (Figure 4.1). Time to publication was measured from the month the abstract was presented at the conference to the month of full publication of its results. The month of publication was not clear for nine (7%) of the full publications and these were excluded from the time to publication analysis. The analysis of time to publication was therefore based on 200 of the 209 trials.

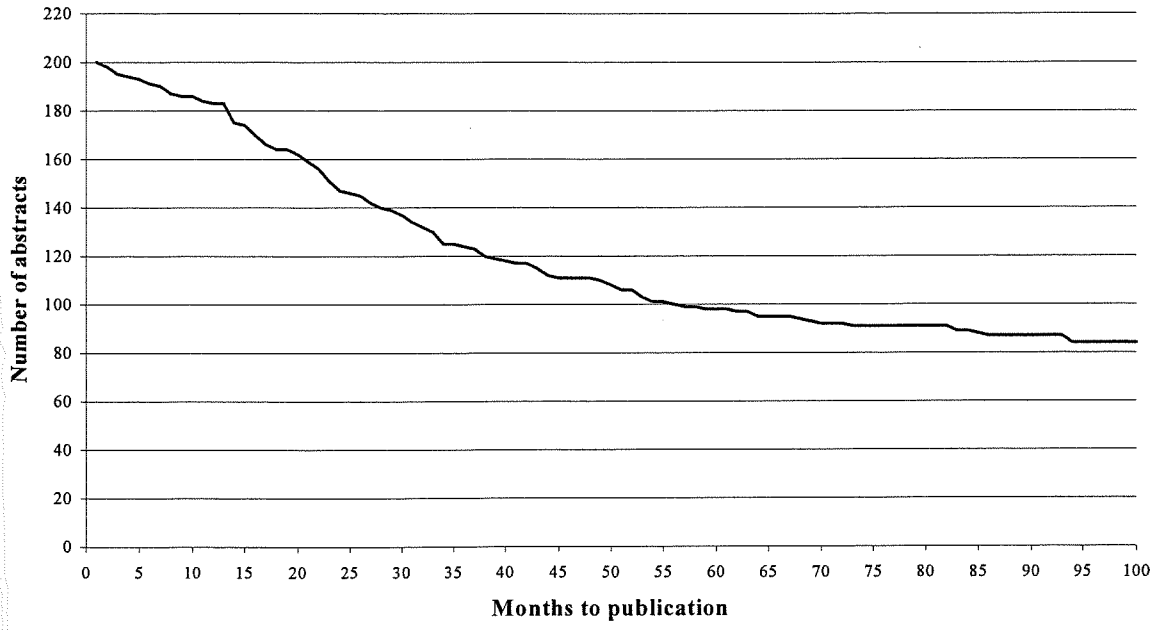


Figure 4:1: Time to publication

Trial characteristics

Two hundred of the 209 abstracts included information on the number of participants randomized. The total number of participants included in these trials was 45,885, ranging between six and 3378 participants per trial. The median number of participants per trial was 120 (IQR 58-256). Those trials reported in conference abstracts that had subsequently been published as a full report were significantly larger than those trials published only as conference abstracts (median 147 (IQR 64-287) for trials published in full versus median 85 (IQR 40-176) for trials published only as abstracts; p=0.0033 using the Mann-Whitley U test) (Table 4.5). The actual method of allocation concealment could be determined in only two of the trials reported in the conference abstracts. It was not possible to determine the method of allocation concealment used in the majority of the trials because not enough information was provided in the abstract.

Table 4:5: Characteristics of the trials reported as abstracts at the American Society of Clinical Oncology conference

	Overall	**Published in full**	**Published as abstracts**
Total number of abstracts	209	125	84
Total number of participants*	45,885	30,133	15,752
Median number of participants (IQR)	120 (IQR 58-256)	147 (IQR 64-287)	85 (IQR 40-176)
Range of participants	6-3,478	8-3,478	6-2,030
Allocation concealment	Adequate = 2 Unclear = 207	Adequate = 0 Unclear = 125	Adequate = 2 Unclear = 82

*Data for the number of participants are given for 200 abstracts, nine abstracts did not include this information (six published and three unpublished).

Identifying possible systematic reviews

No possibly relevant Cochrane review was found for 186 of the 209 trials reported in the conference abstracts (Table 4.6). Twenty-three of the 209 trials reported in

the abstracts were already included in a Cochrane review; 19 reviews referenced the trial as it had been reported in the full publication and four referenced the trial as reported in the conference abstract. Eleven of the 23 trials included in a Cochrane review were referenced as an "Included Study"; eight of these referenced the full publication and three referenced the abstract. Eight of the trials were referenced as an "Excluded Study"; of these, seven referenced the full publication and one referenced the abstract. Four trials were referenced as an "Additional Reference"; all four of these reviews referenced the full publication. In addition, three of the 209 trials reported in the abstracts were included in Cochrane reviews using their individual patient data and cited as an "Included Study".

For two of the 209 trials reported in the conference abstracts, it was unclear if they met the full eligibility criteria for the review and the author of each review was contacted for clarification. Both trial reports were said not to be eligible for inclusion by the authors of the reviews.

Table 4:6: Trials reported as abstracts at the American Society of Clinical Oncology conference for which there was a possible Cochrane review

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Trial already in a Cochrane review:	23	19	4
cited as an included study	11	8	3
cited as an excluded study	8	7	1
cited as an additional reference	4	4	0
No review identified	186	106	80

Discussion

A large number of trials were identified from handsearching conference abstracts published in the *Australian and New Zealand Journal of Medicine* and in the proceedings of the American Society of Clinical Oncology conference in 1992. In both studies, based on the searches that were conducted, approximately 60% of the trials reported in these abstracts were subsequently published as full reports, on average within two to three years of being presented at the conference. The remaining research was not published in full and therefore could not easily have been identified by means other than the abstract or expert knowledge. This figure is consistent with other studies showing that just over half of all conference abstracts describing the results of clinical trials are subsequently published in full (Scherer and Langenberg, 2004). If the search had been broadened to include other bibliographic databases, it is possible that more full reports would have been identified, although this number is likely to have been small. It is also possible that a small number of trials published as conference abstracts in the *Australian and New Zealand Journal of Medicine* in the year 2000 will have been subject to a time-lag before full publication which would mean that their full publications were not available at the time of searching. However, this would not have substantially affected the results of this study. These findings therefore support The Cochrane Collaboration's international effort of handsearching often inaccessible material for reports of controlled trials to enable those who carry out systematic reviews to identify as complete a set of relevant studies as possible through a single electronic resource (Dickersin et al, 2002).

Searching the abstracts from the annual conferences of the Thoracic Society of Australia and New Zealand and the American Society of Clinical Oncology identified a number of reports of randomized trials, which were potentially eligible for inclusion in systematic reviews of healthcare interventions. Twenty-four of the 187 trials, reported in conference abstracts from the Thoracic Society of Australia and New Zealand, were already included in Cochrane reviews. However, the majority of the reviews referenced the trial as reported in the full publication. Cochrane reviews are updated every two years, it is therefore plausible that a trial reported in one of the abstracts was originally included in a Cochrane review and updated data from it was used instead when the full publication became available. If this were to occur, the trial abstract should still be included in the list of references and this was not the case in any of the reviews assessed in this study.

Despite rigorous searches, only three trials were identified which were suitable for inclusion but were currently missing from a Cochrane review, all of which were only published in conference abstracts. Furthermore, the reviewers said that they would still need to contact the authors of the trial for further clarification. In comparison, 23 of the 209 trials reported as abstracts from the American Society of Clinical Oncology conference were already cited in Cochrane reviews. Of these, only 11 trials reported in the conference abstracts were eligible for inclusion in the "Included Studies" section of the review. Eight of these reviews referenced the full publication of the trial and therefore superseded the abstract. Only three Cochrane reviews referenced the trial reported in the conference abstract.

The majority of the trials identified were therefore not judged relevant for inclusion in an existing Cochrane review. There are several reasons why this might be so. This research was carried out within two specific areas of health care (cancer and respiratology); it is possible that these results may not be comparable in other areas of health care. It might also be possible that many of the trials presented in the conference abstracts described new and innovative ideas that were not relevant to current Cochrane reviews. It could also be that Cochrane reviews tend to focus on a specific question and as such may not be broad enough to encompass the breadth of knowledge currently available. However, the most likely explanation is probably that there are currently too few Cochrane reviews to deal with the topics covered in the abstracts, as a recent study predicts that at least another 10,000 Cochrane reviews are needed to cover a substantial proportion of the trials that have already been identified in health care (Mallett and Clarke, 2002).

This study did not find that there was a large number of trials reported in conference abstracts, which were potentially eligible, but not yet included in Cochrane reviews. However, it did identify a number of issues in the quality of the reporting of trials published in conference abstracts. For example, the method of allocation concealment was unclear and poorly reported in the majority of the trials reported in the conference abstracts. The limited amount of information reported about a trial was also a problem in trying to identify whether it was relevant for inclusion in a systematic review. Furthermore, several of the trials did not include data on the number of participants actually included in the trial. This problem of the limited amount of information available about a trial reported in a conference abstract is routinely faced by those carrying out systematic reviews who often need

to contact the authors of the trial for additional information (Clarke and Greaves, 1996). The amount of information reported and the quality of the information reported in trials presented as conference abstracts is, therefore, worthy of more detailed investigation.

Limitations of the existing evidence

A large number of trials were identified which were reported only in conference abstracts and which had not subsequently been published in full. The majority of the trials, however, were not relevant for inclusion in an existing Cochrane review. The results of the study presented in this chapter identified a number of practical issues in trying to include trials reported only in conference abstracts in systematic reviews. These issues will be investigated further in the following two chapters of this thesis. In summary:

- Further evidence is needed to understand better how well trials published as conference abstracts are reported and whether this has improved over time.

- Further evidence is needed to increase understanding of issues around the methodological quality of trials reported in conference abstracts.

- Further evidence is needed to investigate whether there is a difference in the quality and reporting of trials reported as conference abstracts only compared to those that are subsequently published in full.

CHAPTER 5: QUALITY OF TRIALS REPORTED IN THE GREY LITERATURE

Introduction

This thesis has identified that a major source of grey literature in systematic reviews of healthcare interventions are the results of randomized trials which are submitted, presented at scientific meetings and summarized as abstracts in the meeting proceedings. Only just over half of these trials presented as abstracts at scientific meetings are subsequently published in full (Scherer and Langenberg, 2004) and failure to publish is strongly linked to the significance of the results of the trial (Song et al, 2000; Bardy, 1998; Dickersin, 1997; Wormald et al, 1997; Dickersin and Min, 1993; Dickersin et al, 1992; Easterbrook et al, 1991).

This means that for a large proportion of trials, the only public domain source of information about a trial and its results is the abstract published in the conference proceedings. It is important that these abstracts should contain sufficient information about the trial, its methodology and results, to allow an independent judgement of its quality and to be a lasting record of its results. However, chapter four showed that trials reported in conference abstracts may lack important information. In addition, several studies have highlighted that abstracts submitted to scientific meetings may have inadequacies in their reporting of the aims, methods, results and conclusions of the study or, indeed, lack numerical or statistical data (McIntosh, 1996; Panush et al, 1989). The study described in this chapter aims to understand better how well trials found in the grey literature (in particular those reported as conference abstracts) are reported and whether this has improved over

time. The results of this study have been presented at the 11th International Cochrane Colloquium in Barcelona, Spain, in 2003 (Hopewell and Clarke, 2003) and have also been submitted for publication.

Objectives

To assess how well trials presented in the grey literature are reported and whether the quality of the reporting of these trials has improved over time.

The first part of this study aimed to assess whether there is a systematic difference in the quality of trials reported in the grey literature compared to those which are subsequently published in full. The second part of this study aimed to assess whether the quality of trials reported in the grey literature has improved over time.

Source of data

In part one of this study, the sample of grey literature comprised conference abstracts containing reports of randomized trials and published in the proceedings (oral and poster presentations) of the American Society of Clinical Oncology annual conference in 1992. In part two, the sample of grey literature was expanded to include conference abstracts containing reports of randomized trials and published in the proceedings of the American Society of Clinical Oncology annual conferences in 1992 and 2002.

Methods

Handsearching

All conference abstracts published in the proceedings of the American Society of Clinical Oncology annual conferences in 1992 and 2002 were read to identify reports of randomized trials, based on The Cochrane Collaboration's definition of a randomized controlled trial. As described in chapter four of this thesis, this definition requires that "the participants (or other units) were definitely or possibly assigned to one of two (or more) alternative forms of health care using a process of random allocation".

Assessing the quality of the abstract

A checklist was devised to assess the quality of the reporting of trials published in the proceedings of the American Society of Clinical Oncology conferences (Figure 5.1). The checklist was adapted from existing quality criteria such as the CONSORT statement for improving the quality of reports of randomized trials published in journals (Moher et al, 2001b) and suggestions for the structured reporting of journal abstracts and short reports (Deeks and Altman, 1998; Haynes et al, 1996; Haynes et al, 1990; Mulrow et al, 1988). Information obtained during a workshop in Warwick in 2003 on the quality of trials reported in the grey literature (Hopewell and Eisinga, 2003) was also used in developing this checklist. The workshop was run again in Edinburgh in 2004 (Hopewell and Eisinga, 2004) and provided similar information to the earlier one.

The checklist (Figure 5.1) assessed the following reporting criteria:

Authors: Did the conference abstract provide adequate correspondence details for the authors of the study, which would be sufficient to contact them?

Objectives: Did the conference abstract clearly state the objectives of the trial? Did the abstract identify the trial by name or another form of identifier such as a trial number? Did the conference abstract report the date of the trial?

Study design: Did the title of the conference abstract include the word "randomized" and/or "trial"?

Study quality: The methodological quality of each trial reported in the conference abstract was assessed based on the method of allocation concealment (the process used to prevent foreknowledge of group assignment in the randomized controlled trial) (Schulz and Grimes, 2002a). Each trial reported in the conference abstract was assessed as having either adequate allocation concealment (e.g. central randomization office) or inadequate allocation concealment (e.g. date of birth or case record number) or not stated. The method of blinding was also assessed and whether or not the participants, those administering the intervention and/or those assessing the outcomes were blinded to the group assignment and, if so, how this was done. Each conference abstract was assessed as having been either adequately blinded, inadequately blinded or the method of blinding was not stated. Blinding is important to prevent bias because if those involved in the trial are unaware of the

assigned intervention they will not be influenced by that knowledge (Schulz and Grimes, 2002b).

Participants: Did the conference abstract describe broadly the characteristics of the participants who are eligible to be included in the trial (for example who are the participants and how were they selected)?

Interventions: Did the conference abstract describe what the experimental and comparator intervention(s) were and provide some information about how these were administering and for how long? Were the numbers of participants randomized to each intervention given?

Outcomes: Did the conference abstract describe the trial as closed and, if not, how many participants had been randomized to date and how many participants were included in the analysis? Was the intention-to-treat principle followed (i.e. an analysis of participant outcomes based on the groups in which they were randomized irrespective of whether they actually received the planned intervention) and, if not, were the reasons for exclusions made explicit in the abstract?

Results: Did the conference abstract describe a summary of the outcome in each group (for example, the proportion of participants with or without the event) together with the difference between the groups? Did the abstract present confidence intervals for the difference between groups and, where appropriate, was the level of statistical significance presented?

Conclusions: Did the conference abstract provide a description of the primary conclusion(s) based on the results of the trial?

Criteria	Description	Reported
Authors	Adequate correspondence details	
Objectives	Clearly state study objectives	
	Trial name or identifier given	
	Date of trial given (i.e. any date which fixes the trial in time)	
Study design	"Trial" in the title	
	"Randomized" in the title	
Study quality	Method of allocation concealment	
	adequate	
	inadequate	
	not stated	
	Method of blinding	
	adequate	
	inadequate	
	not stated	
Participants	Broad characteristics of eligible participants described (i.e. inclusion and/or exclusion criteria)	
Interventions	Experimental intervention described	
	Comparator intervention(s) described	
	Number of participants randomized (eligible) to experimental intervention described	
	Number of participants randomized (eligible) to comparator intervention(s) described	
Outcomes	Does the abstract describe the trial as closed?	
	yes	
	ongoing	
	not clear	
	Number of participants randomized described	
	Number of participants analysed described	
	Intention-to-treat analysis stated (or reason for exclusion given)	
	Adverse (unintended) events described	
Results	Summary result for each group and its precision (e.g. 95% CI) and, where appropriate, the level of statistical significance	
Conclusions	Primary conclusion(s) described	

Figure 5:1: Checklist for reporting a randomized trial as an abstract

Data extraction and analysis

Data were extracted from each of the 1992 and 2002 conference abstracts containing a report of a randomized trial for each of the quality criteria described in Figure 5.1. Information was collated using a Microsoft Excel spreadsheet and then analysed using a statistical software package (Stats Direct version 2.2.3) as described below.

In part one of this study, the quality of trials reported as conference abstracts in 1992 were analysed and each trial was assessed to determine if it had been published as a full report (as described in chapter four of this thesis). The quality of trials reported as conference abstracts in 1992, which had subsequently been published in full, was then compared to the quality of those that remained unpublished. This was based only on the information reported about each trial in its conference abstract. In the second part of this study, the quality of trials reported as conference abstracts in 1992, irrespective of whether they had subsequently been published in full, were compared to the quality of trials reported as conference abstracts in 2002. In both studies, this was done by comparing the frequency of an event occurring in one set of abstracts with the frequency of an event occurring in the other set. Statistical significance was assessed using the Yates Corrected Chi-Square test or the Fisher-Freeman-Halton Exact test (where cells were less than 5). The difference between medians in both groups was analysed using the Mann-Whitley U test.

Results

Part 1: Quality of published and unpublished trials reported in abstracts

The first part of this study compared the quality of trials reported as conference abstracts, which had subsequently been published in full, with those which remained unpublished based on the information reported in the conference abstract. A total of 209 reports of randomized trials were identified from searching the conference proceedings of the American Society of Clinical Oncology in 1992. Of these, 125 trials initially presented as abstracts had subsequently been published in full with a median time to publication of 27 months. Further details on the number of trials published, the types of cancer and the time to publication have been described in Chapter Four.

Reporting of authors

In 1992, 99% (n=207) of the trials reported in the conference abstracts provided adequate correspondence details for the author(s) to enable someone who found the abstract to contact the trialists for additional information (Table 5.1).

Reporting of objectives

The majority of the 209 trials reported in the conference abstracts clearly stated the objectives of the study or the major hypothesis to be tested.

Reporting of trial name or identifier

Only 14% (n=29) of the trials reported in the conference abstracts described the trial name or gave a trial identifier. This is important as it alerts a person finding more than one report for a particular trial to possible duplicates. These figures were similar for those conference abstracts that had subsequently been published in full and those which remained unpublished (Table 5.1).

Reporting of date of trial

Forty-four percent (n=93) of the trials reported in the conference abstracts reported the date of the trial. This could be any date that fixed the trial in time (for example the start of accrual of the trial or its completion date). This proportion was slightly higher for trials that had subsequently been published in full (48%) compared to those which remained unpublished (38%), but this difference was not statistically significant (p=0.1268) (Table 5.1).

Table 5:1: Reporting of author correspondence details, objectives and study design

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Adequate correspondence details	207 (99%)	124 (99%)	83 (99%)
Clearly state study objectives	207 (99%)	123 (98%)	84 (100%)
Trial name or identifier given	29 (14%)	18 (14%)	11 (13%)
Date of trial given	93 (44%)	61 (48%)	32 (38%)
"Trial" in the title	75 (36%)	51 (41%)	24 (29%)
"Randomized" in the title	76 (36%)	50 (40%)	26 (31%)

Reporting of study design

Overall, only 36% of the trials reported in the conference abstracts included the word "trial" (n=75) or the word "randomized" (n=76) in the title of the abstract.

This number was slightly higher for those trials reported as abstracts that were subsequently published in full (Table 5.1). It has been suggested that authors should routinely put "RCT" at the end of the title or, that the producers of the meeting proceedings could do so automatically if they had a box on the abstract form for authors to tick, should they be reporting a randomized trial (Clarke and Greaves, 1996). This would then allow people to identify randomized trials more easily.

Reporting of study quality

Only two trials reported adequate methods of allocation concealment (e.g. central randomization). None reported clearly inadequate methods. The method of allocation concealment could not be determined for the majority (99%) of the trials reported in the conference abstracts because of a lack of available information. The reporting of the method of blinding was only slightly better, 10% (n=20) of trials reported in the conference abstracts described adequate blinding (e.g. double blind) and four trials described inadequate methods of blinding (e.g. open label).

Table 5:2: Reporting of study quality

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Method of allocation concealment:			
adequate	2 (1%)	0 (0%)	2 (2%)
not stated	207 (99%)	125 (100%)	82 (98%)
Method of blinding:			
adequate	20 (10%)	12 (10%)	8 (9%)
inadequate	4 (2%)	2 (2%)	2 (2%)
not stated	185 (88%)	111 (88%)	74 (88%)

For the majority of trials (88%) reported in the abstracts it was not possible to determine the method of blinding due to an insufficient amount of information. There was no difference in what was reported between those trials that had

subsequently been published in full and those that remained unpublished (Table 5.2).

Reporting of participants

Ninety-three percent (n=195) of the trials reported in the conference abstracts broadly described the characteristics of the participants who were eligible for inclusion in the trial. For example, the abstract described who the participants were, the type of cancer under investigation and the stage of the cancer. This information was unclear for the remaining 14 abstracts (Table 5.3).

Table 5:3: Reporting of participants and interventions

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Characteristics of eligible participants were described	195 (93%)	118 (94%)	77 (92%)
Experimental intervention was described	207 (99%)	124 (99%)	83 (99%)
Comparator intervention(s) was described	201 (96%)	122 (98%)	79 (94%)
Number of participants randomized (eligible) to experimental intervention was described	91 (43%)	55 (44%)	36 (43%)
cross-over trial (number given)	4 (2%)	2 (2%)	2 (2%)
multi-arm trial (number per group given)	16 (8%)	10 (8%)	6 (7%)
number of participants not given	98 (47%)	58 (46%)	40 (48%)
Number of participants randomized (eligible) to comparator intervention(s) was described	90 (43%)	55 (44%)	35 (42%)
cross-over trial (number given)	4 (2%)	2 (2%)	2 (2%)
multi-arm trial (number per group given)	16 (8%)	10 (8%)	6 (7%)
number of participants not given	99 (47%)	58 (46%)	41 (49%)

Reporting of interventions

The majority of the trials reported in the conference abstracts described clearly what the experimental (99%) and comparator (96%) intervention(s) were and provided some information about how they were administered and for how long. The number of participants randomized to the experimental and comparator intervention(s) was described in 53% of the trials. Of these, four were reports of

cross-over trials and 16 were reports of multi-arm trials and the numbers of participants randomized to each group were described. This means that just under half (47%) of the trials reported in the conference abstracts did not describe the number of participants randomized to either the experimental or the comparator intervention(s). This figure was the same for those trials that had subsequently been published in full and those that remained unpublished (Table 5.3).

Reporting of outcomes

Reporting of trial status

Only 7% (n=14) of the trials reported in the conference abstracts were described as closed to accrual and the specified period of follow-up had been completed. Thirty-three percent (n=69) of the trials were described as ongoing and for the remaining 60% (n=126), it was not clear if the trial was closed or whether it was still ongoing (Table 5.5). Again, there was no difference between those trials subsequently published in full and those which remained unpublished.

Reporting of number of participants randomized

Encouragingly, the majority (96%) of the trials reported in the conference abstracts included information on the total number of participants randomized. This information was unclear for nine of the 209 conference abstracts. Overall, the median number of participants randomized per trial was 120 (IQR 58-256) (Table 5.4). This number was significantly larger (p=0.0033) for those trials reported in conference abstracts that were subsequently published in full compared to those which remained unpublished. The median number of participants per trial was 147

(IQR 64-287) for trials subsequently published in full compared to 85 (IQR 40-176) for those that remained unpublished.

Reporting of number of participants analysed

The number of participants included in the analysis was less well reported; only 67% (n=139) of trials reported in the conference abstracts included information on the total number of participants analysed. This information was unclear or not available in approximately one-third (n=70) of the conference abstracts. Overall, the median number of participants included in the analysis per trial was 85 (IQR 46-190) (Table 5.4). This number was again higher for those trials that were subsequently published in full (median 99; IQR 51-262) than for those which remained unpublished (median 69; IQR 32-119) (p=0.0057).

Table 5:4: Reporting of number of participants randomized and analysed

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Total number of participants randomized was described	200 (96%)	119 (95%)	81 (96%)
Number of participants randomized	45,885	30,133	15,752
Median number randomized	120	147	85
Inter-quartile range	IQR 58-256	IQR 64-287	IQR 40-176
Range	6-3478	8-3478	6-2030
Total number of participants analysed was described	139 (67%)	83 (42%)	56 (67%)
Number of participants analysed	20,819	15,605	5214
Median number analysed	85	99	69
Inter-quartile range	IQR 46-190	IQR 51-262	IQR 32-119
Range	6-1516	10-1516	6-458

Reporting of intention-to-treat analysis

Twenty-eight percent (n=59) of the trials reported in the conference abstracts indicated, or suggested, that they had followed the intention-to-treat principle and

analysed participants based on the groups into which they were randomized irrespective of whether they actually received the planned intervention. Eighteen (9%) trials clearly did not follow the intention-to-treat principle but gave reasons why participants who had been randomized to a specific intervention were not included in the overall analysis. However, for over half (63%) of the trials it was not clear from the conference abstract whether the intention-to-treat principle had, or had not, been followed. This was similar for trials, which were subsequently published in full and those that remained unpublished (Table 5.5).

Table 5:5: Reporting of trial status and intention- to-treat analysis

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Trial was described as:			
closed	14 (7%)	9 (7%)	5 (6%)
ongoing	69 (33%)	42 (34%)	27 (32%)
not clear	126 (60%)	74 (59%)	52 (62%)
Intention to treat analysis:			
yes	59 (28%)	34 (27%)	25 (30%)
reason for exclusion given	18 (9%)	9 (7%)	9 (11%)
not clear	132 (63%)	82 (66%)	50 (59%)

Reporting of results

Twenty-seven percent of the trials reported the size of the treatment effect and its statistical significance. For example this included the number of events in each group and a p-value for the difference or the percentage in each group and a p-value. Thirty-four percent of the trials only reported the size of the treatment effect (for example the number of events in each group or the percentage in each group) and 17% of the trials only reported the statistical significance of the treatment effect. The remaining conference abstracts (22%) reported none of the above (Table 5.6). Of these abstracts, the most common results to be presented were an overall

percentage for all participants combined. Five conference abstracts did not contain any results but said that these would be presented at the meeting.

Table 5:6: Reporting of results

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Type of results reported:			
size of effect	72 (34%)	42 (34%)	30 (36%)
size of effect and significance	56 (27%)	37 (29%)	19 (23%)
significance of effect	36 (17%)	27 (22%)	9 (11%)
other	45 (22%)	19 (15%)	26 (31%)

The statistical significance of the results of each trial reported in conference abstracts was assessed based on the results presented for the primary outcome measure. All of the trials assessed were cancer trials and therefore the primary outcome measure was generally survival, the response rate, the time to disease progression or the disease recurrence rate. Statistical significance was assessed based on a trial's reported p-value. The trialist may have reported an actual p-value or that a result was significant at a particular level (i.e. $p<0.05$ or $p<0.01$). The number of trials reporting statistically significant results was significantly greater for those trials subsequently published in full (44% n=54) compared to those trials which remained unpublished (21% n=17) ($p<0.012$). Fourteen percent (n=11) of trials, which remained unpublished, reported statistically non-significant results compared to 9% (n=11) of trials that were subsequently published in full. Attempts were also made to determine the direction of the results of the trial (i.e. the result was significant in favour of the experimental arm of the trial or significant in favour of the comparator arm of the trial). However, it was not always possible to determine which was the experimental arm of the trial or indeed both arms of the

trial might have been experimental interventions, without more knowledge and information about the intervention under investigation; this part of the analysis was therefore abandoned.

Statistical significance was not always reported in the conference abstract. In this case clinical significance was assessed. A result was considered to be clinically significant if the wording in the abstract suggested that it clearly favoured a particular arm of the trial; for example the abstract included wording such as "there was a significant difference between ... " or "there was a statistically significant difference...". Eleven percent (n=9) of trials which remained unpublished and 5% (n=6) of trials that were subsequently published in full reported clinical significance using this definition. In addition, 32% (n=26) of trials which remained unpublished and 25% (n=31) of trials that were subsequently published in full reported clinical insignificance (i.e. reported that "there was no significant difference..."). It should be recognised, however, that this is a very subjective and crude way to measure significance, as it is entirely dependent on the wording in the conference abstracts. It does, however, give an indication as to how and whether the trialists themselves perceived the importance of the trial results. Finally, for 22% (n=18) of trials which remained unpublished and 17% (n=21) of trials that were subsequently published in full, it was not possible from reading the conference abstract to determine either the statistical or clinical significance of the results using this basic assessment criteria. These figures are based on the 204 trials for which results were reported in the conference abstracts.

Reporting of conclusions

The majority (n=203) of the trials reported in the conference abstracts provided details of primary conclusions. Six of the trials reported that it was too early to provide any such conclusions (Table 5.7). There was no difference in the reporting of conclusions between those trials that were subsequently published in full and those which remained unpublished.

Table 5:7: Reporting of conclusions

	Overall (n=209)	**Published (n=125)**	**Unpublished (n=84)**
Primary conclusion(s) was described:			
yes	203 (97%)	124 (99%)	79 (94%)
none given	6 (3%)	1 (1%)	5 (6%)

Part 2: Quality of trials reported in abstracts over time

The above assessment of the quality of trials reported in abstracts at the American Society of Clinical Oncology conference in 1992 showed that some aspects of the trials were well reported and some were not. Interestingly, there appeared to be little significant difference in reporting between those trials that were subsequently published in full and those trials which remained unpublished. The exception to this rule was the size of the trial and the statistical significance of the trial results. Those trials subsequently published in full tended to be larger and more likely to have statistically significant results than those which remained unpublished. In the second part of this study, the quality of trials reported in conference abstracts was assessed to see whether the quality of the reporting had improved over time. The quality of trials presented in the proceedings of the American Society of Clinical

Oncology conference in 1992 (irrespective of whether they were subsequently published in full) was compared to those presented at the same conference ten years later in 2002.

A total of 494 reports of randomized trials were identified from searching the abstracts for the American Society of Clinical Oncology meetings in 1992 and 2002. In 1992, 209 (14%) of the 1471 abstracts presented at the meeting contained reports of randomized trials and, in 2002, 285 (16%) of the 1811 abstracts contained reports of randomized trials. In addition, some abstracts were identified which mentioned a trial which might have been randomized but the results were not the main focus of the abstract. These were not assessed further in this study.

Table 5:8: Number of trials reported as abstracts and site of cancer

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Breast cancer	46 (22%)	58 (20%)
Central nervous system tumours	-	7 (2%)
Gastrointestinal cancer	28 (13%)	44 (15%)
Genitourinary cancer	17 (8%)	29 (10%)
Gynaecologic cancer	11 (5%)	17 (6%)
Head and neck cancer	6 (3%)	12 (4%)
Leukaemia adult	12 (6%)	4 (1%)
Leukaemia paediatric	5 (2%)	2 (1%)
Lung cancer	22 (11%)	56 (20%)
Lymphoma adult	11 (5%)	8 (3%)
Lymphoma paediatric	4 (2%)	3 (1%)
Melanoma	4 (2%)	7 (2%)
Multiple cancers	6 (3%)	5 (2%)
Non specific cancer*	30 (14%)	22 (8%)
Paediatric solid tumours	4 (2%)	3 (1%)
Sarcoma	3 (1%)	8 (3%)

*Examples of non-specific cancer included trials of pain management, anti-emetics and quality of life.

Trials were identified in a number of different cancer specialities. The most common categories were breast cancer (22% in 1992 and 20% in 2002), gastrointestinal cancer (13% in 1992 and 15% in 2002), genitourinary cancer (8% in 1992 and 10% in 2002) and lung cancer (11% in 1992 and 20% in 2002) (Table 5.8).

Reporting of authors

In 1992 and 2002, 99% of the trials reported in the conference abstracts provided adequate correspondence details for the author(s) of the trial (Table 5.9).

Table 5:9: Reporting of author correspondence details, objectives and study design

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Adequate correspondence details	207 (99%)	284 (99%)
Inadequate correspondence details	2 (1%)	1 (1%)
Clearly state study objectives	207 (99%)	285 (100%)
Trial name or identifier given	29 (14%)	72 (25%)
Date of trial given	93 (44%)	128 (45%)
"Trial" in the title	75 (36%)	131 (46%)
"Randomized" in the title	76 (36%)	153 (54%)

Reporting of objectives

In both years, 99% of the abstracts clearly stated the objectives of the study or the major hypothesis to be tested.

Reporting of trial name or identifier

In 1992, only 14% (n=29) of the abstracts reported the trial name or a trial identifier, this improved to 25% (n=72) in 2002 (p=0.0028) but the proportion is still low.

Reporting of date of trial

Forty-five percent of all conference abstracts reported the date of the trial.

Reporting of study design

Only 36% (n=75) of the trials in 1992 included the word "trial" in the title of the conference abstract; this increased to 46% (n=131) in 2002. In 1992, 36% of the abstracts included the word "randomized" in the title and again this improved in 2002 to 54% (n=153) (p=0.0002) (Table 5.9).

Reporting of study quality

In both years, only one to two percent of trials reported in the conference abstracts stated the method of allocation concealment. For the remaining trials, the method of allocation concealment could not be determined because of a lack of information in the conference abstract. The method of blinding was slightly better reported. In 1992, 10% (n=20) of the trials reported in the conference abstracts described adequate blinding (e.g. "double blind"), this increased to 14% (n=40) in 2002 (Table 5.10).

Table 5:10: Reporting of study quality

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Method of allocation concealment:		
adequate	2 (1%)	5 (2%)
not stated	207 (99%)	280 (98%)
Method of blinding:		
adequate	20 (10%)	40 (14%)
inadequate	4 (2%)	7 (2%)
not stated	185 (88%)	238 (84%)

In both years, two percent of trials described lack of blinding (e.g. "open label"). However, for the majority of trials, it was not possible to determine the method of blinding due to insufficient information.

Reporting of participants

In 2002, more conference abstracts described the characteristics of the participants who were eligible for inclusion in the trial: 93% (n=195) in 1992 compared to 97% (n=281) in 2002 (p=0.0027) (Table 5.11).

Table 5:11: Reporting of participants and interventions

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Characteristics of eligible participants described	195 (93%)	281 (97%)
Experimental intervention described	207 (99%)	277 (97%)
Comparator intervention described	201 (96%)	277 (97%)
Number of participants randomized (eligible) to experimental intervention described	91 (43%)	128 (45%)
cross-over trial (number given)	4 (2%)	8 (3%)
multi-arm trial (number per group given)	16 (8%)	25 (9%)
number of participants not given	98 (47%)	124 (44%)
Number of participants randomized (eligible) to comparator intervention described	90 (43%)	128 (45%)
cross-over trial (number given)	4 (2%)	8 (3%)
multi-arm trial (number per group given)	16 (8%)	25 (9%)
number of participants not given	99 (47%)	124 (44%)

Reporting of interventions

In both years, approximately 98% of the trials reported in the conference abstracts described what the experimental and comparator intervention(s) were and provided some information about how they were administered and for how long. This information was not clear for the few remaining abstracts. The number of participants randomized to the experimental and comparator intervention(s) were given in 53% (n=111) of trials in 1992 and 56% (n=161) of trials in 2002 (Table

5.11). Just under half of the abstracts did not describe the number of participants randomized to either the experimental or comparator intervention(s).

Reporting of outcomes

Reporting of trial status

Seven percent (n=14) in 1992 and 10% (n=27) in 2002 of the trials reported in the conference abstracts were described as closed and the intended period of follow-up had been completed. A third of the trials (33%) in both years, were described as ongoing and for just over half (60% in 1992 and 56% in 2002) it was not clear if the trial was closed or whether it was still ongoing (Table 5.13).

Reporting of number of participants randomized

Overall, 96% of the trials reported in the conference abstracts included information on the total number of participants randomized. This was unclear for nine abstracts in 1992 and 15 abstracts in 2002. The median number of participants randomized per trial was significantly higher ($p<0.001$) in 2002 (median 209; IQR 96-418) compared to 1992 (median 120; IQR 58-256) (Table 5.12).

Reporting of number of participants analysed

The number of participants included in the analysis was less well reported. In 1992, 67% (n=139) of conference abstracts reported the number of participants analysed; this dropped significantly to 49% (n=139) in 2002 ($p=0.0001$). Part of the explanation for this drop is that 29 (10%) of the 285 trials in 2002 did not contain any results, compared to only five in 1992. However, even when abstracts that did not contain any results were excluded from the analysis, this significant difference

in reporting of the number of participants included in the analysis was still present; 67% (n=136) in 1992 compared to 52% (n=132) in 2002 (p=0.0015).

Table 5:12: Reporting of number of participants randomized and analysed

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Total number of participants randomized was described	200 (96%)	270 (95%)
Number of participants randomized	45885	245293
Median number randomized	120	209
Inter-quartile range	IQR 58-256	IQR 96-418
Range	6-3478	12-154000
Total number of participants analysed was described	139 (67%)	139 (49%)
Number of participants analysed	20819	40060
Median number analysed	85	124
Inter-quartile range	IQR 46-190	IQR 58-236
Range	6-1516	10-3507

There was also a difference in the median number of participants included in the analysis compared to the number of participants initially randomized. In 1992, the median number of participants analysed per trial was 85 (IQR 46-190) compared to 124 (IQR 58-236) in 2002 (p=0.0065). In comparison, the median number of participants randomized, for these trials alone, was 95 (IQR 51-195) in 1992 and 167 (IQR 82-279) in 2002.

Reporting of intention-to-treat analysis

In 1992, 28% (n=59) of the trials reported in the conference abstracts indicated, or suggested, that they had followed the intention-to-treat principle, compared to only 15% (n=43) in 2002 (p=0.004). Nine percent (n=18) of trials in 1992 and 6% (n=17) in 2002 clearly did not follow the intention-to-treat principle but gave reasons why participants, who had been randomized, were not included in the overall analysis. For the majority of trials, however, it was not clear from the

abstract whether the intention-to-treat principle had or had not been followed (Table 5.13). Again, if those abstracts which did not contain any results were excluded, there was no important change in these results.

Table 5:13: Reporting of trial status and intention to treat analysis

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
The trial was described as:		
closed	14 (7%)	27 (10%)
ongoing	69 (33%)	98 (34%)
not clear	126 (60%)	160 (56%)
Intention to treat analysis:		
yes	59 (28%)	43 (15%)
reason for exclusion given	18 (9%)	17 (6%)
not clear	132 (63%)	225 (79%)

Reporting of results

Twenty-seven percent (n=56) in 1992 and 26% (n=73) in 2002 of the trials reported the size of the treatment effect and its significance (i.e. the difference in the treatment effect between the two groups and its statistical significance). Examples of what was reported included the number of events in each group and a p-value for the difference, the percentage in each group and a p-value, the relative risk and a p-value, and the hazard ratio and a p-value. Thirty four percent (n=72) and 35% (n=101) of the trials only reported the size of the treatment effect; for example the number of events in each group, the percentage in each group, the median in each group and the odds ratio. Seventeen percent (n=36) in 1992 and 12% (n=35) in 2002 of the trials only reported the statistical significance of the treatment effect (Table 5.14).

The remaining conference abstracts, 22% (n=45) in 1992 and 27% (n=76) in 2002, reported none of the above. Of these abstracts, the most common results to be presented were an overall percentage for all participants combined. In addition, 29 conference abstracts in 2002 did not contain any results but said that these would be presented at the meeting. This was the case for five of the 1992 abstracts.

Table 5:14: Reporting of results

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Type of results reported:		
size of effect	72 (34%)	101 (35%)
size of effect and significance	56 (27%)	73 (26%)
significance of effect	36 (17%)	35 (12%)
other	45 (22%)	76 (27%)

The statistical significance of the results of trials reported in the conference abstracts in 2002 were then compared to those from 1992. While it is not possible to assess for evidence of publication bias in the 2002 abstracts, as not enough time has passed since they were presented at the meeting, these findings can be used in a future study to assess the extent to which the results of a trial affect its future publication. The criteria described for part one of this study was used and a similar proportion of trials in 2002 reported statistically significant results (i.e. $p<0.05$) compared with those reported in 1992 (35% (n=71) in 1992 compared to 30% (n=77) in 2002). Eleven percent (n=22) of trials in 1992 and 13% (n=32) of trials in 2002 reported statistically non-significant results (i.e. $p>0.05$). Where a p-value was not reported, clinical significance based on the wording of the abstract was assessed. For 7% (n=15) of trials in 1992 and 16% (n=40) of trials in 2002, the wording in the abstract clearly suggested that the results of the trial clearly favoured

a particular arm of the trial (i.e. the abstract included words such as "there was a significant difference between..."). In addition, 28% (n=57) of trials in 1992 and 18% (n=47) of trials in 2002 described clinically non significant results (i.e. the abstract included words such as "there was no significant difference..."). For the remaining trials (19% (n=39) in 1992 and 23% (n=60) in 2002), it was not possible from reading the conference abstract to determine either statistical or clinical significance without more information and knowledge of the intervention and condition under investigation. It is important to note that these figures are based only on those trials which reported the results of the trial in the conference abstract; as noted above, five abstracts in 1992 and 29 abstracts in 2002 did not report any results but said that these would be presented at the meeting.

Reporting of conclusions

Ninety-seven percent (n=203) in 1992 and 92% (n=263) in 2002 of the trials reported in the conference abstracts provided details of primary conclusions (Table 5.15). The remaining abstracts reported that it was too early in the trial to provide any conclusions.

Table 5:15: Reporting of conclusions

	ASCO 1992 (n=209)	**ASCO 2002 (n=285)**
Primary conclusion(s) described:		
yes	203 (97%)	263 (92%)
none given	6 (3%)	22 (8%)

Discussion

Handsearching the conference proceedings of the American Society of Clinical Oncology identified a large number of abstracts containing reports of randomized trials in a variety of cancer specialities. The percentage of abstracts containing reports of randomized trials was similar in 1992 and 2002, but the size of the trials had increased significantly in 2002. These findings are consistent with those of Chlebowski and Lillington (1994) who searched American Society of Clinical Oncology conference abstracts from 1984 to 1993 to identify breast cancer trials. They found that 17% of abstracts contained reports of randomized trials and that these had increased in size over time. This increase in the size of trials in recent years has also been found in other studies. McDonald and colleagues (2002) investigated trends in the number and size of trials reported in general healthcare journals from 1948 to 1997. They found a continuing increase in the number of larger trials reported over time.

In 1992, those trials reported in abstracts which were subsequently published in full, were significantly larger and more likely to report statistically significant results than those that remained unpublished. Scherer and Langenberg (2004), in a recent Cochrane methodology review, identified that trials with larger sample sizes were more likely to have statistically significant results and were more likely to be published in full. This is perhaps not surprising as trials with small sample sizes are more likely to be underpowered and therefore less likely to achieve statistical significance in tests comparing study treatments. In this situation, non-publication of the results of a trial could be because of a lack of significant results and not necessarily because of the size of the trial. In all cases, trials with small sample

sizes should still be published because, in a systematic review, they still contribute to the available evidence and can be combined statistically in a meta-analysis.

Certain aspects of the trials reported in the 1992 and 2002 conference abstracts were well described and there were no significant differences in the quality of the reporting between those trials from the 1992 meeting, which had subsequently been published in full, and those that remained unpublished. In 1992 and 2002, criteria that were well reported included the objectives of the trial, the characteristics of the participants who were included in the trial and details of the experimental and comparator interventions. However, other areas were poorly reported such as the method of allocation concealment and blinding, the number of participants randomized to each group of the trial, whether the results of the trial were analysed according to the intention-to-treat principle and whether the trial was ongoing or closed. The results of this study support those of others, which also found considerable variation in the quality of abstracts submitted to scientific meetings (McIntosh, 1996; Panush et al, 1989).

What is of particular concern is the reporting of the trial results and the difference in the number of participants included in the analysis compared to the number of participants initially randomized. While this difference was significant in the 1992 conference abstracts, the difference was much larger in 2002. In 2002, half of all trials failed to report the number of participants included in the analysis. Reasons for this are unclear; one possible explanation could be the desire to publish research findings sooner - before the intended period of follow-up is completed and all participants are eligible to be included in the analysis. This has important

implications especially if the results of the trial are not then published in full (Scherer and Langenberg, 2004). An increasing number of trials reported in the 2002 conference abstracts stated that the results of the trial would be presented at the meeting. Again, this could be indicative of a growing trend to present the results of trials sooner at scientific meetings, possibly even before the true effects of the intervention are known.

There is a need to improve the quality of trials reported in the proceedings of scientific meetings. Abstracts submitted to scientific meetings are often subject to peer review, which should provide some level of scientific control. However, even when abstracts submitted to scientific meetings are subject to peer review there is still considerable variation in the evaluation of these abstracts (Timmer et al, 2003; Von Elm et al, 2003; Montgomery et al, 2002; Smith et al, 2002; Vilstrup and Sorensen, 1998; Usatoff and Waxman, 1997; Cho and Bero, 1996; Jacks and Peterson, 1994; Rubin et al, 1993; Corry, 1990). For example, Montgomery and colleagues (2002) investigated the agreement between referees who peer reviewed abstracts, which had been submitted to a primary care research conference. Despite using a checklist to assess the quality of these studies, there was considerable disagreement between the reviewers as to which studies should be accepted for presentation and which quality components were most important. Clearer guidelines are needed for authors on submitting an abstract for presentation at a scientific meeting such as providing specific instructions that outline the key elements of the study to be described.

Most of the work on improving the quality of summary reports for trials has concentrated on the development of structured abstracts in journal articles. Recommendations from an Ad Hoc Working Group in 1987 (Mulrow et al, 1988), modified in 1990 (Haynes et al, 1990), proposed that abstracts in journals should describe a study using specified headings rather than paragraph format. A number of medical journals such as the *BMJ, Lancet, JAMA, Annals of Internal Medicine*, and the *New England Journal of Medicine* now require authors to provide structured abstracts for articles of clinical interest. Structured abstracts require authors to state systematically the objective, basic research design, clinical setting, participants, interventions (if any), main outcome measurements, results and conclusions of the study. A properly constructed abstract should help professionals quickly assess the reliability and content of the report and also aid the indexing and retrieval of reports from bibliographic databases such as MEDLINE (Harbourt et al, 1995).

Structured abstracts in journal articles face similar limitations to those of abstracts submitted to scientific meetings, the most obvious being the limited amount of detail on the study's methodology and results that authors can include. The average word limit for a structured abstract submitted to a scientific journal such as the *BMJ* or *JAMA* is 300 words. In comparison, the average word length for abstracts published in the proceedings of the American Society of Clinical Oncology conference in 2002 was 250 to 300 words.

Despite some journal requirements that the abstracts of clinical studies should be presented in a structured format, the quality of these abstracts can still be poor and

the structured format does not necessarily lead to an improvement in the quality of what is reported (Dupuy et al, 2003; Pitkin et al, 1999; Scherer and Crawley, 1998; Froom and Froom, 1993; Narine et al, 1991). Scherer and Crawley (1998) compared the quality of trials published in two specialised ophthalmology journals before they introduced the use of structured abstracts and in the year following their introduction. The CONSORT statement was used to evaluate the quality of these reports. However, in Scherer and Crawley's study the introduction of structured abstracts was not associated with an improvement in reporting quality. In comparison, Taddio and colleagues (1994) compared the quality of structured and unstructured abstracts of research articles in three large general medical journals and found that structured abstracts tended to be much better reported.

Limitations of the existing evidence

There is a need to improve the quality of the reporting of trials submitted to scientific meetings by providing authors with specific instructions on the key elements of the study to be described. Meanwhile, questions have also been raised about the reliability of abstracts submitted to scientific meetings and possible differences between these abstracts and their subsequent full paper (Chokkalingham et al, 1998; Weintraub, 1987). Research is needed to address these important issues and ultimately to help improve the quality of trials presented in and reported at scientific meetings.

- Further evidence is needed to assess how the content and quality of the information reported for a trial in a conference abstract differs from that in a subsequent full publication.

CHAPTER 6: COMPARING TRIALS REPORTED IN THE GREY LITERATURE WITH THEIR FULL PUBLICATION

Introduction

The study described in Chapter Five illustrated a number of concerns over the poor reporting and quality of trials published in the proceedings of scientific meetings (Hopewell and Clarke, 2003; McIntosh, 1996; Panush et al, 1989). One of the major concerns was a lack of adequate information about key aspects of the trial and how representative the results presented in the abstract might be, compared to those subsequently published in a full journal article. This chapter aims to build on these findings to assess how the information reported for a trial in a conference abstract differs from that reported in a subsequent full publication. The results of this study have been presented at the 25th Annual Meeting of the Society of Clinical Trials in New Orleans, USA, in 2004, for which it was awarded a Thomas Chalmers Student Scholarship Program award (Hopewell, 2004). This study has also been accepted for presentation at the 12th International Cochrane Colloquium in Ottawa, Canada, in October 2004.

Objectives

To assess whether the content and quality of the information reported for a trial in the grey literature (specifically, in a conference abstract) differs from that reported in its subsequent full publication.

This study is divided into two parts, the first compares whether the information reported for a trial in a conference abstract differs from that reported in the text of its subsequent full publication. The second part takes this process a step further and compares whether the information reported for a trial in a conference abstract differs from that reported in the abstract of the identified subsequent full publication. This is an important comparison because decisions about the suitability of a report for a systematic review might be based on the abstract of a full article, or the reviewers might not be able to obtain the full article when conducting their review.

Methods

As has been described in Chapter Four, all conference abstracts published in the proceedings of the American Society of Clinical Oncology annual conference in 1992 were searched by hand to identify reports of randomized trials. Subsequent full publications were identified for each trial reported in the conference abstract by searching *The Cochrane Central Register of Controlled Trials (CENTRAL)* in *The Cochrane Library* (Issue 4, 2002) for the first author (and then second author or third if this was not successful) listed in the abstract. The search was repeated using the U.S. National Library of Medicine's PubMed database (December 2002) for those conference abstracts for which a full report was not found in *CENTRAL*. Possible publications were examined to see if the full report corresponded with the subject matter in the conference abstract. If multiple publications were identified for the trial then the full publication that most closely corresponded to the conference abstract was selected.

Source of data

A total of 209 reports of randomized trials were identified from searching the proceedings of the American Society of Clinical Oncology annual conference in 1992. Full publications were found for 125 of the 209 abstracts and these covered a wide variety of cancer specialities. For the purposes of this study a sample of 40 trials were selected from these 125 conference abstracts for further examination. This sample was selected so as to include all trials in three specific areas of cancer: gastrointestinal cancer (n=17), genitourinary cancer (n=11) and lung cancer (n=12).

Data extraction

The information on the trial reported in the original conference abstract was compared to the information in its subsequent full publication by a team of experienced systematic reviewers. This team consisted of individuals who were experienced in the methodology of systematic reviews and who had previously expressed interest in this area of research. These interested individuals were contacted first by email and then by a letter describing the proposed project in detail and asking them if they would take part in the study.

Two reviewers independently read each conference abstract and full publication and extracted relevant data. One reviewer (Sally Hopewell) read and extracted data on each of the 40 abstracts and full publications to ensure overall consistency in the data extraction. Data were extracted from each conference abstract and corresponding full publication using a standardised data extraction form (Appendix Three), which was developed based on the experiences of developing the checklist to assess the quality of trials reported as conference abstracts as described in

Chapter Five. Data were extracted on the year of publication, the number of authors and type of authorship (e.g. individual or group authors), the title, the length of the report, the source of funding and ethical approval.

Information was also extracted on correspondence details, the objectives of the study, the trial name and date, the study design, the study quality (e.g. the method of allocation concealment and the method of blinding), the type of participants included in the trial, the interventions assessed and the number of participants randomized. Data were extracted on the primary outcome assessed, whether the trial was closed, the number of participants included in the analysis, if the analysis was carried out according to the intention-to-treat principle, if any adverse events were reported, what results were reported and what conclusions were given. Any changes between what was reported in the conference abstract and in the full publication were recorded for each of the above criteria.

Three trials were excluded during the data extraction stage. Two trials were excluded because the full publication relating to the abstract was not published in English; one was published in Italian and one in Japanese. One trial was excluded because it was found not to be a randomized trial when the full publication was read (Figure 6.1).

Data analysis

Data extracted from each conference abstract and its corresponding full publication were entered into an Excel spreadsheet and checked electronically for consistency between the two reviewers. Data were analysed using a statistical software package

(STATA version 8.0, Stata Corporation, Texas, USA). For each of the data criteria extracted, the difference between what was reported in the abstract and what was reported in the full publication was tabulated. Statistical significance was assessed using the Chi-Square test or the Fisher Freeman Halton exact test (where cells <5). The mean difference in the number of participants randomized in each group was analysed using a paired t test. A single estimate (%) of agreement between what was reported in the abstract and in the full publication was also calculated.

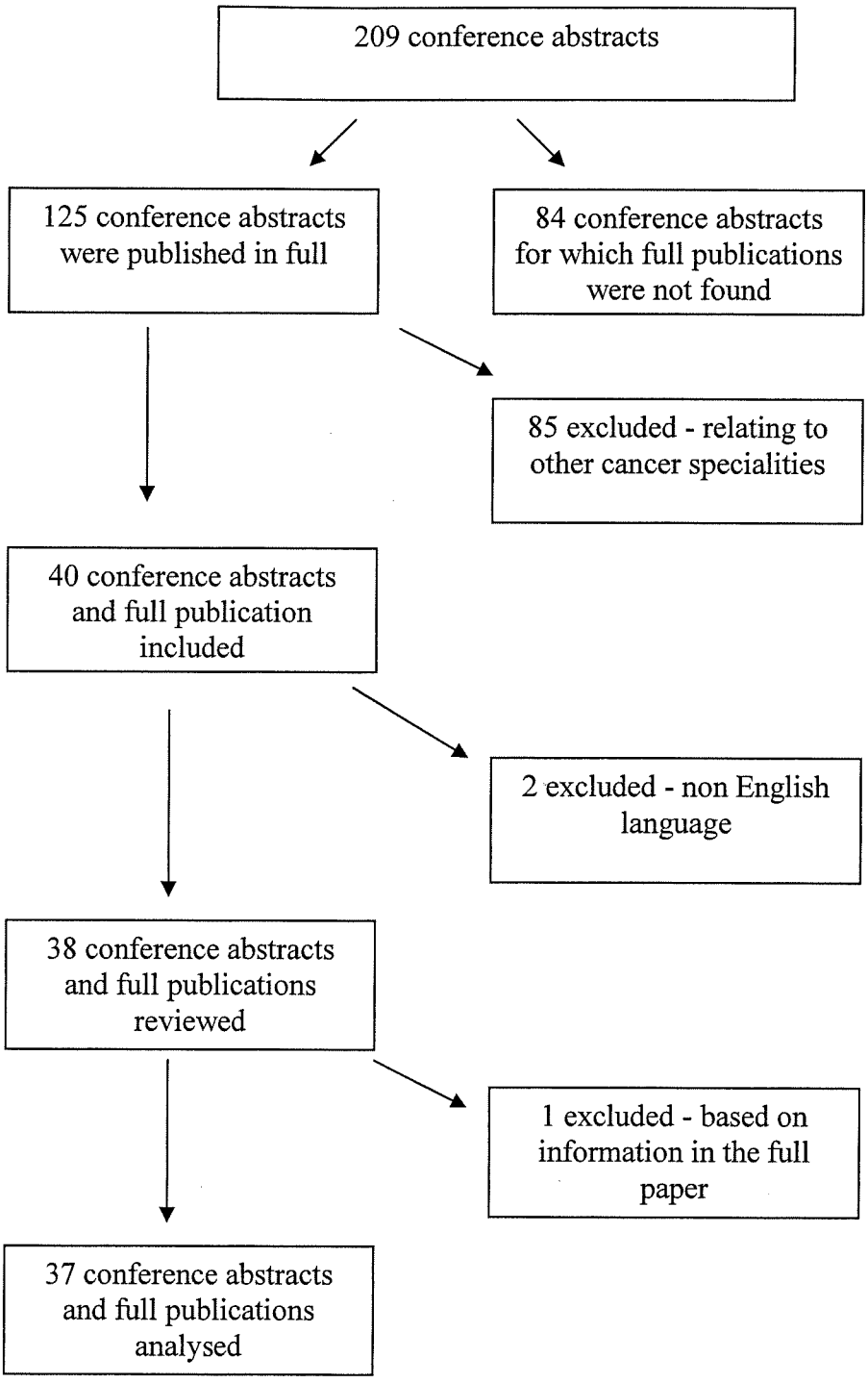


Figure 6:1: Flow diagram showing the study eligibility criteria

Results

Part 1: Comparing trials reported as conference abstracts with their subsequent full publication

The first part of this study compared whether the content and quality of information reported for a trial in a conference abstract differs from that reported in its subsequent full publication.

Reporting of general characteristics

Time to publication

Of the 37 trials assessed, the median time to publication from presentation at the conference to full publication of results was 31 months (IQR 20-51), ranging from one month to seven years (Figure 6.2).

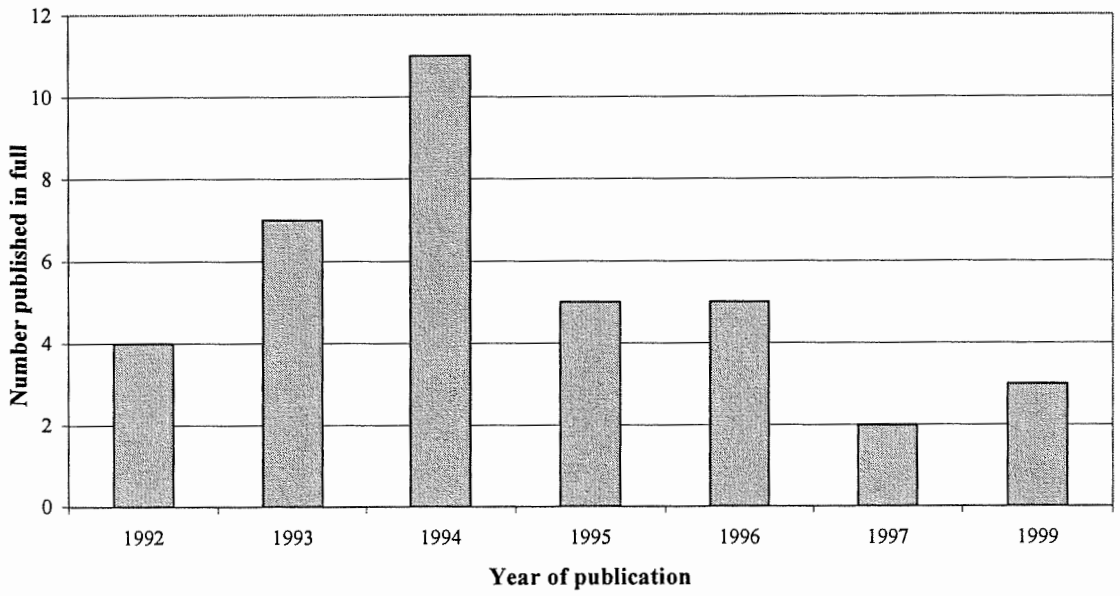


Figure 6:2: Time to full publication

This figure is based on the month the abstract was presented at the conference (March 1992) to the month the trial was published as a full journal article.

Authorship

Thirty-four trials (92%), in both the conference abstract and corresponding full publication, had a list of named individuals as authors and two trials had corporate authors (i.e. a study group name, rather than individual authors). The remaining trial gave a single author and the study group name in the conference abstract and group authorship in the full publication.

Changes in authorship between the conference abstract and the full publication were common. Only one trial had the same authors, in the same order, in both the abstract and the full publication. Three trials had the same authors in both but they appeared in a different order. The most common change between the abstract and the full publication was the addition of new authors (Figure 6.3). Thirty-one trials had between one and 11 new authors listed on the full publication; the median number of new authors per trial was three (IQR 2-5). Two trials had fewer authors in the full publication. Thirty-one of the 37 trials had the same first author in both the conference abstract and the full publication.

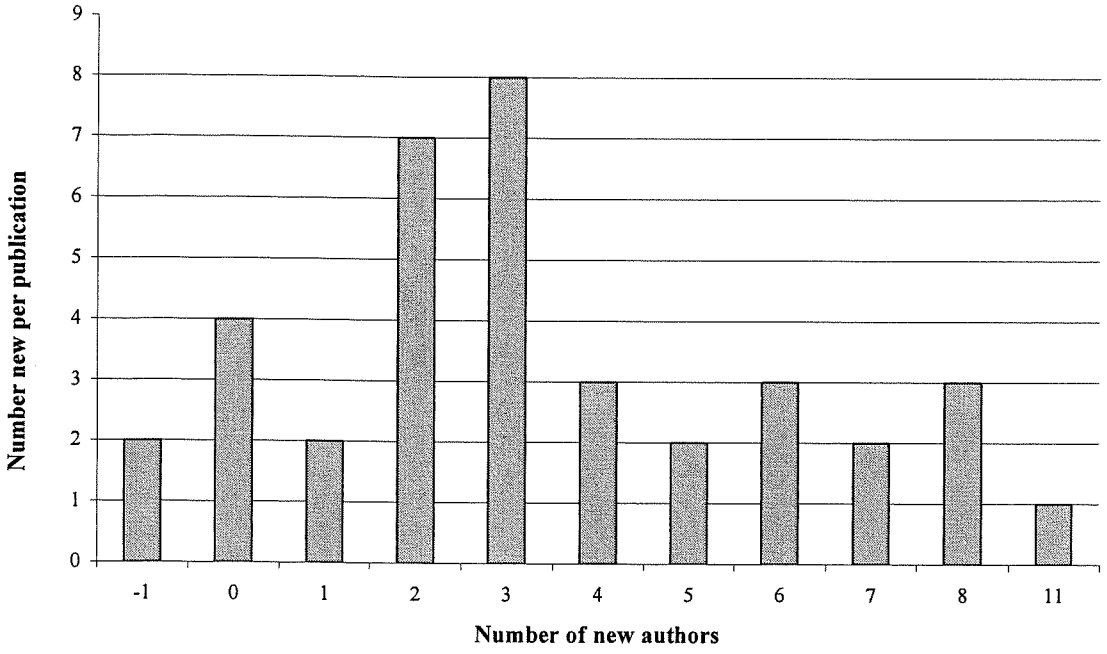


Figure 6:3: Number of new authors listed in the full publication

Title

The title for seven (19%) of the 37 conference abstracts was the same as its full publication. There was a minor change to the title in 20 (54%) conference abstracts and their corresponding full publication. For example the order of the wording had changed in the full publication (six trials), abbreviations in the conference abstract had been written in full in the full publication (three trials) and the wording was slightly different in the full publication but the meaning remained the same (11 trials).

There was a major change to the title for the remaining 10 (27%) trials. These included a shift of emphasis in the full publication compared to the conference abstract (four trials) (i.e. the abstract was declarative whereas the full publication

was neutral, or vice versa), more specific information was given in the abstract and the full publication was more general (two trials), and more specific information was given in the full publication and the abstract was more general (four trials).

Multi-centre status

For the majority of trials, information on multi-centre status was only available from the full publication. Thirty-four (90%) trials were multi-centre trials, one was a single-centre trial and, for three trials, it was not clear if they were single- or multi-centre trials. This information was not available for any of the trials reported in the conference abstracts.

Length of report

The average length of a full publication was six pages (IQR 5-7), ranging from three to 12 pages per publication. In comparison, the average length of a conference abstract was approximately 250 to 300 words.

Funding source

Information on the trial's source of funding was not given in any of the conference abstracts. From the full publications, eight of the 37 trials (22%) were funded by charitable grants, eight (22%) were funded through multiple funding sources, four (11%) were funded by industry and three trials (8%) were funded by the government. Fourteen (37%) trials did not describe their sources of funding (Figure 6.4).

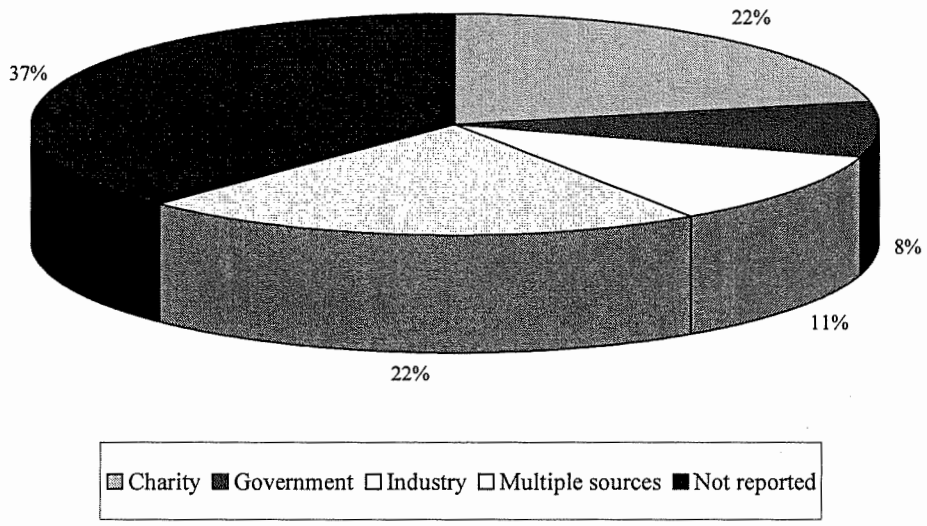


Figure 6:4: Source of funding

Ethical approval

Details of ethical approval for the trial were also only available in some of the full publications and never in the conference abstracts. For ten trials (27%), ethical approval was obtained either through the local ethics committee or the institutional review board; for seven trials (19%) ethical approval was not described in the full publication but was implied as patients gave informed consent before entering the trial. For the remaining 20 trials (54%), no details of ethical approval or consent were given.

Reporting of authors

Thirty-six (97%) trials reported in the conference abstracts and full publications provided adequate correspondence details to enable someone to contact the trialists

for additional information. Of these, 34 (92%) provided the same contact details (more detail was usually given in the full publication) and two gave different contact details. One trial only gave adequate correspondence details in the full publication and not in the conference abstract.

Reporting of objectives

Thirty-five (95%) trials clearly stated the objectives of the study in both the conference abstract and full publication and, as would be expected, more detail was provided in the full publication. The remaining two trials did not clearly report the objectives of the study in the conference abstracts but did so in the full publication.

Reporting of trial name or identifier

Eleven trials (30%) reported the trial name or identifier in both the conference abstract and the full publication. There was no change in the reporting of the trial name or identifier for 32 (86%) of the 37 trials. For three trials, the trial identifier was reported in the conference abstract only and for two trials in the full publication only (Table 6.1).

Table 6:1: Reporting of trial name or identifier

Abstract	**Paper**		
	No	**Yes**	**Total**
No	21	2	23
Yes	3	11	14
Total	24	13	37

Reporting of date of trial

Nineteen trials reported the date of the trial in both the full publication and conference abstract. This date was the same for 16 (43%) of the 19 trials. However, for three trials the start and/or finish date were different in the two sources. Thirteen trials reported the date of the trial in the full publication only and two trials in the conference abstract only (Table 6.2). Overall, the difference between the number of trials reporting the date of the trial in the full publication compared to the abstract was statistically significant (p=0.0099).

Table 6:2: Reporting of date of trial

Abstract	Paper		
	No	Yes	Total
No	3	13	16
Yes	2	19	21
Total	5	32	37

Reporting of study design

Eleven trials (30%) included the word "trial" in the title of both the full publication and the conference abstract. Seven trials included the word "trial" in the title of the full publication only and two in the conference abstract only (Table 6.3).

Table 6:3: Reporting of "trial" in title

Abstract	Paper		
	No	Yes	Total
No	17	7	24
Yes	2	11	13
Total	19	18	37

Eighteen trials (49%) included the word "randomized" in the title of both the full publication and the conference abstract. Four trials included the word "randomized" in the title of the full publication only and two in the conference abstract only (Table 6.4).

Table 6:4: Reporting of "randomized" in title

Abstract	Paper		
	No	Yes	Total
No	13	4	17
Yes	2	18	20
Total	15	22	37

Reporting of study quality

Seventeen trials (46%) reported the method of allocation concealment in the full publication (Table 6.5). This was not reported in any of the conference abstracts ($p<0.0001$). The method of allocation concealment included central randomization by computer (ten trials), central randomization by telephone (five trials), and sealed numbered envelopes held centrally (two trials).

Table 6:5: Reporting of method of allocation concealment

Abstract	Paper		
	No	Yes	Total
No	20	17	37
Yes	0	0	0
Total	20	17	37

No difference could be detected in the quality of reporting of allocation concealment following the introduction of the CONSORT statement in 1996 (Begg et al, 1996). Fifteen of the 32 trials published between 1992 and 1996 described

adequate allocation concealment and two of the five trials published between 1997 and 1999 described adequate allocation concealment.

The method of blinding was poorly reported. Only seven full publications and six conference abstracts reported adequate methods of blinding of the study participants. One trial reported a lack of blinding in the full publication (e.g. open label). The vast majority of trials (78%) did not report on blinding in either the conference abstract or the full publication (Table 6.6). It is recognised that it can be difficult to blind participants in some of the trials investigated here but, even for these trials, no information was reported for the blinding of outcome assessors, which may have been possible.

Table 6:6: Reporting of method of blinding

Abstract	Paper			
	Adequate	**Inadequate**	**Not stated**	**Total**
Adequate	6	0	0	6
Not stated	1	1	29	31
Total	7	1	29	37

Reporting of participants

The majority of conference abstracts and full publications (92%) described the type of participants eligible for inclusion in the trial. More information was available in the full publication; for example on the trial's inclusion and exclusion criteria, the participants' age, sex, stage of cancer and prognosis. The conference abstracts for three trials did not report the type of participants included in the trial.

Reporting of interventions

All of the trials (100%) described the experimental and comparator interventions in both the conference abstracts and full publications. Again, the full publication provided more information on how the interventions were administered (e.g. dose, route and duration). More participants were reported in the full publication compared to the conference abstract for 12 trials in regard to the experimental intervention and 11 trials in regard to the comparator intervention; for eight of these, the conference abstract explicitly reported interim results only. Fewer participants were reported in the full publication for three trials in the experimental and two in the comparator interventions. The same number of participants randomized was reported in the full publication and conference abstract for 11 trials in the experimental and 13 in the comparator interventions. Finally, 11 trials reported the number of participants randomized to each arm of the trial in the full publication only, and not in the abstract (Table 6.7).

Table 6:7: Reporting of intervention groups

Type of change reported	Experimental intervention (n=37)	Comparator intervention (n=37)
Number of participants in full paper only	11	11
More participants in full paper	4	3
More participants in full paper (abstract interim report only)	8	8
Fewer participants in full paper	3	2
Same number of participants reported	11	13

Reporting of outcomes

All 37 trials reported cancer-related outcome measures. For the vast majority of trials, the primary outcome measures were either survival, time to disease progression or disease recurrence rates. The same primary outcome was reported in

the conference abstracts and full publications for 31 (84%) of the 37 trials. Again, more detail was available in the full publication. For five trials, the primary outcome was different in the full publication compared to the conference abstract and for one trial, the primary outcome was reported only in the full publication.

Reporting of trial status

Thirty-two trials (86%) were described as closed and the intended period of follow-up had been completed in the full publication; this was the case for only seven (19%) of the trials reported in the conference abstracts. Fourteen trials reported in the conference abstracts were still ongoing and for 16 it was not clear if the intended period of follow-up had been completed (Table 6.8). This difference between what was reported in the full publication compared to the abstract was statistically significant ($p<0.001$).

Table 6:8: Reporting of trial status

Abstract	Paper			
	Closed	**Ongoing**	**Not clear**	**Total**
Closed	7	0	0	7
Ongoing	14	0	0	14
Not clear	11	0	5	16
Total	32	0	5	37

Reporting of number of participants randomized

All of the trials reported in the full publications included information on the total number of participants randomized. This was not clear for three trials reported in the conference abstracts. The median number of participants randomized per trial was 120 (IQR 82-217) in the full publications and 117 (IQR 72-233) in the

conference abstracts; this difference was not statistically significant (p=0.2202) (Table 6.9).

In the 34 trials that reported the number of participants randomized in the abstract and the full publication, six trials (18%) reported an increase of 10% or more in the total number of participants randomized in the full publication. Three trials reported a decrease of at least 10% in the total number of participants randomized in the full publication compared to the conference abstract.

The total numbers of participants randomized were the same for 17 (46%) of the 37 conference abstracts and full publications. Twelve full publications reported more participants compared to the conference abstracts, six of these abstracts reported interim analyses only and one failed to report an abandoned arm of a trial, which was reported in the full publication. Five trials described fewer participants randomized in the full publications compared to the conference abstracts (Table 6.10).

Table 6:9: Reporting of number of participants randomized and analysed

	Abstract (n=37)	**Paper (n=37)**
Number of participants randomized described	34	37
Total number of participants randomized	6008	6406
Median number randomized (IQR)	117 (72-233)	120 (82-217)
Range	8-612	8-612
Number of participants analysed described	25	36
Total number of participants analysed	3004	6013
Median number analysed (IQR)	96 (53-139)	117 (73-213)
Range	8-462	6-612

Reporting of number of participants analysed

The number of participants included in the analysis was less well reported. Thirty-six of the full publications gave information on the number of participants analysed in the trial, compared to only 25 of the conference abstracts. One full publication did not give the number of participants analysed; this trial reported pre- and post-interferon antibodies levels in renal cell carcinoma. For the 25 trials that reported the number of participants analysed in the abstract and full publication, the median number analysed per trial was 96 (IQR 53-139) in the conference abstracts compared to 117 (IQR 73-213) in the full publications (Table 6.9). This difference in the number of participants included in the analysis was statistically significant (p=0.003).

In the 25 trials that reported the number of participants analysed in the abstract and the full publication, 12 (48%) trials reported an increase of at least 10% in the total number of participants analysed in the full publication. One trial reported more than 10% fewer participants analysed in the full publication.

The numbers of participants included in the analysis were the same for only eight (22%) of the 37 conference abstracts and full publications. Fourteen full publications reported more participants included in the analysis compared to the conference abstracts; ten of these stated that they reported interim analyses only. The full publications for four trials described fewer participants analysed compared to the conference abstracts (Table 6.10).

Table 6:10: Changes in the number of participants reported

Type of change reported	Participants randomized (n=37)	Participants analysed (n=37)
Number of participants in full paper only	3	11
More participants in full paper	6	4
More participants in full paper (abstract interim analysis only)	6	10
Fewer participants in full paper	5	4
Same number of participants reported	17	8

Reporting of intention to treat

Seventeen trials (46%) indicated, or suggested, that they had followed the intention-to-treat principle in the full publications compared to only five (14%) in the conference abstracts (Table 6.11). Ten trials reported in the full publications and one trial reported in the conference abstract did not follow the intention-to-treat principle but gave reasons why participants, who had been randomized, were not included in the overall analysis. For the majority of trials reported in the conference abstracts (84%), it was not clear if the intention-to-treat principle had been followed because of a limited amount of information. This difference in the level of reporting between the abstract and the full publication was statistically significant ($p<0.0001$).

Table 6:11: Reporting of intention-to-treat analysis

Abstract	Paper			
	Yes	Reason given	Not clear	Total
Yes	4	1	0	5
Reason given	0	1	0	1
Not clear	13	8	10	31
Total	17	10	10	37

Reporting of adverse effects

Thirty-three (89%) full publications and 26 (70%) conference abstracts included details of adverse or unintended effects of the treatment (Table 6.12). All of the full publications gave a greater level of detail. Seven full publications reported adverse effects, which were not described in the conference abstract (p=0.0806).

Table 6:12: Reporting of adverse effects

Abstract	Paper		
	No	Yes	Total
No	4	7	11
Yes	0	26	26
Total	4	33	37

Reporting of results

Only four trials (11%) reported similar results in the conference abstract and the full publication. For 11 (30%) trials it was difficult to assess if they gave the same conclusions because limited data were available in the conference abstracts.

The remaining 22 (59%) trials reported different results in the conference abstract and the full publication. This was primarily because the abstract presented preliminary trial results and / or because the length of follow-up was longer in the full publication and, therefore, different numbers of events were included in the analysis (17 trials). Other possible explanations included fewer participants in the full publication (four trials) and because the focus of the trial had changed (one trial).

Given that the type of results presented were rarely the same in both the abstract and its corresponding full publication, the type of summary statistic presented was also different. The majority of trials reported in the conference abstracts presented the percentage in each group (nine trials), the p-value showing the overall significance of the result (seven trials), or the median or mean number of events in each group (six trials). The majority of full publications presented a survival analysis (18 trials) for each group and p-value (18 trials) or the number of events in each group (10 trials).

Overall, nine (24%) abstracts and 25 (68%) full publications reported the size of the treatment effect and the significance of its results. Examples of what was reported included the number of events in each group and p-value, survival analysis for each group and p-value and the mean or median number in each group and p-value. Sixteen (43%) abstracts and 12 (32%) full publications reported the size of the treatment effect (for example the number of events in each group or the percentage in each group) and 11 (30%) abstracts only reported the significance of the trial results.

When those trials in which the number of participants included in the analysis were the same in both the abstract and the full publication (n=8) were assessed separately, it was still difficult to determine if the results presented in both were the same. In three trials the length of follow-up was longer in the full publication and in five trials only very limited information was given in the conference abstracts.

Table 6:13: Type of result reported

	Abstract (n=37)	Paper (n=37)
Type of results reported:		
size of effect	16 (43%)	12 (32%)
size of effect and significance	9 (24%)	25 (68%)
significance of effect	11 (30%)	-
other	1 (3%)	-

Reporting of conclusions

Twenty-seven trials (73%) reported the same conclusions in both the conference abstract and the full publication; more details were provided in the full publication. Two trials reported the conclusions of the trial in the full publication only and one abstract reported that it was too early to give conclusions.

Six trials reported different conclusions in the full publication compared to the conference abstract. Of these, two abstracts said that the treatment was as effective as the control, whereas the full publications said that the treatment was less effective; three abstracts said there was no difference in treatment effect, whereas the full publications said that treatment was effective; one abstract described three treatments whereas the full publication described four. Finally, in one trial the emphasis of the conclusions changed from disease recurrence, in the conference abstract, to reducing infection rates in the full publication.

Part 2: Comparing trials reported as conference abstracts with the abstract of their subsequent full publication

One of the limitations identified in the first part of this chapter was the limited amount of detailed information reported about a trial in a conference abstract when compared with its subsequent full publication. It is possible that the differences identified could be due primarily to the amount of space available in which to report a trial in a conference abstract. However, it is anticipated that some differences between what is reported in a conference abstract and the subsequent full publication, such as the number of participants included and analysed in the trial, would still be apparent, irrespective of the amount of space available in which to report the results of the trial. To test this hypothesis, the second part of this study aimed to compare whether the content and quality of information reported for a trial in a conference abstract differs from that reported only in the abstract of its subsequent full publication. This is also important because decisions about the suitability of a report for a systematic review might be based on the abstract of a full article, or the reviewer might not be able to obtain the full article when conducting their review.

The methods applied were the same as those described in Part One of this chapter with two experienced systematic reviewers assessing each abstract. However, to avoid potential biases, steps were taken to blind the person extracting data from each abstract as to whether it was from a conference or a journal. All identifying features were removed from both the conference abstract and the abstract of the full publication and the text and font size for each abstract were standardised using

Microsoft Word before data extraction took place. Thirty-six of the 37 trials were included in the analysis. One of the 37 trials was excluded because the full publication did not contain an abstract.

Reporting of objectives

Thirty-three (92%) trials stated the objectives of the study in both the conference abstract and corresponding abstract of the full publication. One trial did not report the objectives of the study in the conference abstract but did so in the abstract of the full publication and two trials did not clearly report the objectives of the study in the abstract of the full publication but did in the conference abstract.

Reporting of trial name or identifier

Only three trials (8%) reported the trial name or identifier in both the conference abstract and corresponding abstract of the full publication (Table 6.14). Ten trials (27%) reported the trial name or identifier in the conference abstract and not in the abstract of the full publication (p=0.040).

Table 6:14: Reporting of trial name or identifier

Abstract	**Paper abstract**		
	No	**Yes**	**Total**
No	23	0	23
Yes	10	3	13
Total	33	3	36

Reporting of date of trial

Seven trials (19%) reported the date of the trial in both the conference abstract and corresponding abstract of the full publication. This date was the same for only three

of the seven trials. Thirteen trials reported the date of the trial in the conference abstract only and two trials in the abstract of the full publication only (Table 6.15).

Table 6:15: Reporting of date of trial

Abstract	Paper abstract		
	No	Yes	Total
No	14	2	16
Yes	13	7	20
Total	27	9	36

Reporting of study quality

The method of allocation concealment was not reported in any of the conference abstracts and only one trial reported the method of allocation concealment in the abstract of the full publication. The method of blinding was also poorly reported. Only six of the 36 conference abstracts and corresponding abstracts of the full publications reported on the method of blinding of the study participants.

Reporting of participants

The majority (94%) of conference abstracts and corresponding abstracts of the full publications provided some details of the type of participants eligible for inclusion in the trial.

Reporting of interventions

All of the trials (100%) described the experimental and comparator interventions in both the conference abstracts and corresponding abstracts of the full publications. Twenty-five trials (69%) reported in the conference abstracts described the number

of participants randomized to the experimental and comparator arms of the trial compared to 18 trials (50%) reported in the abstracts of the full publications.

More participants were reported in the abstracts of the full publications for four of these 18 trials and one reported fewer participants compared to the corresponding conference abstracts. The same number of participants randomized was reported in both the conference abstract and abstract of the full publication of only six of the 36 trials (Table 6.16).

Table 6:16: Reporting of intervention groups

Type of change reported	Experimental intervention (n=36)	Comparator intervention (n=36)
Number of participants in abstract of full paper only	7	7
Number of participants in conference abstract only	12	12
More participants in abstract of full paper	4	4
Fewer participants in abstract of full paper	1	1
Number of participants not reported	6	6
Same number of participants reported	6	6

Reporting of outcomes

The same primary outcome was reported in the conference abstracts and corresponding abstracts of the full publications for 32 (89%) of the 36 trials. However, for four trials the primary outcome was different in the abstract of the full publication compared to the conference abstract.

Reporting of trial status

Twenty trials (56%) were described as closed and the intended period of follow-up had been completed in the abstract of the full publication; this was the case for only six (17%) of the trials reported in the conference abstracts. Fourteen trials reported

in the conference abstracts were still ongoing and for 16 conference abstracts and 14 abstracts in the full publications it was not clear if the intended period of follow-up had been completed (p=0.159) (Table 6.17).

Table 6:17: Reporting of trial status

Abstract	Paper abstract			
	Closed	Ongoing	Not clear	Total
Closed	6	0	0	6
Ongoing	7	6	1	14
Not clear	7	8	1	16
Total	20	2	14	36

Reporting of number of participants randomized

The majority of the trials included information on the total number of participants randomized. This was not clear for three trials reported in the conference abstracts and one trial reported in the abstract of the full publication. The median number of participants randomized per trial was 120 (IQR 92-222) in the abstracts of the full publications and 115 (IQR 74-221) in the conference abstracts. The numbers of participants randomized were the same for 15 (42%) of the 36 conference abstracts and corresponding abstracts of the full publications. Twelve abstracts of the full publications reported more participants compared to their corresponding conference abstracts (six of these conference abstracts reported interim analyses only) and six trials described fewer participants randomized in the abstracts of the full publications compared to the corresponding conference abstracts.

In the 33 trials that reported the number of participants randomized in the conference abstract and corresponding abstract of the full publication, six trials (18%) reported an increase of 10% or more in the total number of participants

randomized in the abstract of the full publication and three trials reported a decrease of at least 10% in the total number of participants randomized in the abstract of the full publication.

Reporting of number of participants analysed

The number of participants included in the analysis was much more poorly reported. Twenty-five (69%) of the conference abstracts gave information on the number of participants analysed in the trial compared to only 16 (44%) of the abstracts of the full publications. These were the same number of the conference abstracts and corresponding full publications in only 11 of the 36 trials. Of these 11 trials, the median number of participants analysed was 102 (IQR 73-265) for the trials reported in the conference abstracts and 114 (IQR 97-265) for the trials reported in the abstracts of the full publications. The number of participants included in the analysis were the same in only four of the 11 conference abstracts and corresponding abstracts of the full publications; five trials reported more participants in the abstract of the full publication and two trials reported fewer participants analysed in the abstract of the full publication.

Reporting of intention-to-treat analysis

Only four (11%) conference abstracts and three (8%) abstracts of the full publications indicated or suggested that the trial had followed the intention-to-treat principle. Three trials, reported in the abstracts of the full publications, and one trial, reported in the conference abstract, said that the intention-to-treat principle had not been followed but gave reasons why participants, who had been randomized, were not included in the overall analysis.

Table 6:18: Reporting of intention-to-treat analysis

Abstract	Paper abstract			
	Yes	Reason given	Not clear	Total
Yes	1	0	3	4
Reason given	0	0	1	1
Not clear	2	3	26	31
Total	3	3	30	36

For the majority of trials it was not clear if the intention-to-treat principle had been followed because of a limited amount of information in both the conference abstract and the abstract of the full publication (Table 6.18).

Reporting of adverse effects

Twenty-five (69%) of the 36 conference abstracts and abstracts of the full publications included some details of adverse or unintended effects of the treatment (Table 6.19).

Table 6:19: Reporting of adverse effects

Abstract	Paper abstract		
	No	Yes	Total
No	10	1	11
Yes	1	24	25
Total	11	25	36

Reporting of results

Only four (11%) of the 36 trials reported similar results in both the conference abstract and the corresponding abstract of the full publication. For 11 (30%) trials it was difficult to assess if they reported the same conclusions because of the limited amount of information available. For the remaining 21 (58%) conference abstracts

and 17 (47%) abstracts of the full publications, different results were reported in each of the corresponding abstracts and four abstracts of the full publications did not describe the results of the trial.

Reporting of conclusions

Twenty-nine (80%) of the 36 trials reported the same conclusions in both the conference abstract and the abstract of the full publication. One trial reported the conclusions of the trial in the abstract of the full publication only and one conference abstract reported that it was too early to give conclusions. Four trials reported different conclusions in the abstract of the full publication compared to the conference abstract. Of these, two conference abstracts said that the treatment was as effective as the control, whereas the abstracts of the full publications said that the treatment was less effective and two conference abstracts said there was no difference in treatment effect, whereas the abstract of the full publications said that the treatment was effective. Finally, in one trial the emphasis of the conclusions changed from disease recurrence, in the conference abstract, to reducing infection rates in the abstract of the full publication.

In the second part of this study comparing the information reported for a trial in a conference abstract with that reported in the abstract of its subsequent full publication, some areas were well reported with a high level of agreement between the two sets of abstracts. These included details of the trial objectives, the characteristics of the participants who were eligible for inclusion in the trial, details of the experimental and comparator interventions, the primary outcome measure and the conclusions. Obviously, when comparing these findings with those

described in the first part of this chapter, there was more detailed information provided about the trial in the full publication than when compared to the abstract of the full publication (Figure 6.20).

Perhaps surprisingly, some aspects of the trial were more poorly reported in the abstract of the full publication than the conference abstract. These included details such as the trial name or identifier, the date of the trial and the number of participants randomized to each arm of the trial. There was a poor level of agreement between conference abstract and abstract of full publication for the number of participants randomized and the number included in the analysis; this difference was also apparent when the entire full publication was assessed. However, it was not anticipated that the number of participants included in the analysis would be so poorly reported in the abstract of the full publication. Only 44% of abstracts of the full publications reported the number of participants analysed compared to 69% of the conference abstracts. Given the inconsistency between the two sets of abstracts, the results for primary outcome were rarely the same. Details of study quality were also poorly reported in both the conference abstracts and abstracts of the full publications. Finally, more trials were described as closed in the abstract of the full publication compared to their corresponding conference abstract and again this reflects the findings from the first part of this study.

Table 6:20: Summary of level of agreement between conference abstracts, full publications and abstracts of the full publications

Criteria assessed		Abstract (n=37)	Paper (n=37)	Level of agreement (%)	Abstract (n=36)	Paper abstract (n=36)	Level of agreement (%)
Authors	Author correspondence details were the same	36	37	97%	-	-	-
Objectives	Study objectives were described	35	37	95%	35	34	92%
	Trial name or identifier given	14	13	30%	13	3	8%
	Date of trial given	21	32	43%	20	9	19%
Study design	"Trial" in the title	13	18	30%	-	-	-
	"Randomized" in the title	20	22	49%	-	-	-
Study quality	Method of allocation concealment was described	0	17	0%	0	1	0%
	Method of blinding was described	6	7	16%	6	6	17%
Participants	Characteristics of eligible participants were described	34	37	92%	34	35	94%
Interventions	Experimental intervention was described	37	37	100%	36	36	100%
	Comparator intervention was described	37	37	100%	36	36	100%
	Participants randomized to experimental intervention were described	26	37	30%	25	18	17%
	Participants randomized to comparator intervention were described	26	37	35%	25	18	17%
Outcomes	Primary outcome measure was described	36	37	84%	35	35	89%
	Trial status was described (e.g. closed)	7	32	19%	6	20	17%
	Number of participants randomized was described	34	37	46%	33	35	42%
	Number of participants analysed was described	25	36	22%	25	16	11%
	Intention-to-treat principle was described	5	17	14%	4	3	11%
	Important adverse effects were described	26	33	70%	25	25	67%
Results	Results for primary outcome were described	37	37	11%	36	32	11%
Conclusions	Primary conclusions were described	34	37	73%	34	36	80%

Discussion

Those who carry out systematic reviews need to decide whether, at the time of doing a systematic review, they should include information about a trial published in a conference abstract, if a fuller publication is not available. In this study, 60% of randomized trials presented as abstracts at the American Society of Clinical Oncology in 1992 had been published in full with a median time to publication of 31 months. This is similar to other studies, which found that around 56% of trials presented as conference abstracts had been published in full, with a median time to publication of 20 months (Scherer and Langenberg, 2003).

Changes in the authorship and title between what was reported in the conference abstract and in the corresponding full publication were relatively common. These findings are supported by other studies, which also found significant changes in the authorship and title of the study between the conference abstract and full publication. In these studies, changes in authorship occurred 30% to 52% of the time (Bhandari et al, 2002; Ciesla and Wojcik, 2001; Roy et al, 2001) and changes in the title occurred 27% to 63% of the time (Sprague et al, 2003; Bhandari et al, 2002; Weintraub, 1987). These changes may be an indication that research was still ongoing at the time of submitting the abstract to the conference and that a significant contribution to the research or manuscript preparation was made after that time.

Some aspects of the trials were well reported and, for these areas, there was generally a high level of consistency between what was reported in the conference abstract and in

the full publication. The most important of these areas included the objectives of the study (95% agreement), the characteristics of the participants eligible for inclusion in the trial (92% agreement), what the experimental and comparator interventions were (100% agreement) and what the primary outcome measure was (84% agreement). These findings are consistent with those of Bhandari and colleagues (2002) who examined the consistency of 159 abstracts, presented at an annual orthopaedic surgeons conference in 1996, with their subsequent full publications. Only seven of the abstracts in the Bhandari study were reports of randomized trials, the rest were reports of observational studies. The study by Bhandari and colleagues (2002) found that study objectives were the same 99% of the time, study design was the same 99% of the time and primary outcome measures were the same 87% of the time.

In other areas, there were major changes between what was reported in the conference abstracts compared to their corresponding full publications. The most concerning of these were the number of participants randomized to experimental and comparator arms of the trial (30-35% agreement), the total number of participants randomized (46% agreement), and the total number of participants analysed (22% agreement). Given this high level of disagreement it was therefore unsurprising that the trial results were rarely consistent between the conference abstract and full publication. The most likely explanation for this considerable variation between the conference abstract and full publication is that many conference abstracts reported the preliminary results of a trial that was still ongoing, either still recruiting participants to the trial or completing the specified period of follow-up. All the trials assessed in this study were cancer trials

and the principal outcome measure was survival. Survival is often a long-term outcome measure and it is therefore not too surprising that the full publication will report longer follow-up than the conference abstract. This difference in the reporting of the number of participants randomized into the trial and included in the analysis was also apparent in the second part of this study which compared conference abstracts with the abstracts of their subsequent full publications. However, here the numbers of participants randomized to each arm of the trial and those included in the analysis were particularly poorly reported in the abstract of the full publication compared to the conference abstract or full article.

In most cases, the changes in the number of participants did not significantly change the conclusions of the trial (73% agreement). However, in a small number of cases the change was large enough so that a treatment was described as effective in the conference abstract and less effective in the full publication, or vice versa. This could have important clinical implications if the conclusions of a specific trial reported as an abstract were used to guide clinical practice. It could also have important implications for systematic reviews, especially for those reviews or meta-analyses with only a small number of included trials and reviews including only trials reported as abstracts. This impact would be most important when the abstracts for these trials reported interim analyses.

Differences in the number of participants included in the full publication compared to the conference abstract have been noted in several studies assessing the rate of full

publication of abstracts. Weintraub (1987), in a study of 33 abstracts presented at surgical meetings, found that 30% of full publications had fewer participants than their corresponding abstract and 27% of full publications had more participants than their corresponding abstract. Despite these changes, the conclusions of the study remained the same 70% of the time. Another study assessed the publication rate of abstracts presented at a perinatology conference in 1997 (Davies et al, 2002). They found that 26% of full publications had an increase in the number of participants and 13% of full publications had a decrease in the number of participants compared to the corresponding conference abstracts; the conclusions remained the same 65% of the time. Klassen and colleagues (2002) conducted a similar study of abstracts presented at the Society of Pediatric Research (1992-1995) and found that 48% of full publications reported an increase in sample size and 11% of full publications reported a decrease in sample size when compared to the corresponding conference abstracts. Only one study has focused specifically on reports of randomized trials and compared the agreement of data in abstracts of randomized trials to data from their subsequent full publications (Chokkalingam et al, 1998). In this study of randomized trials presented as abstracts at an ohpthalmology conference in 1988 to 1989, 33% of the data in the abstract disagreed with the published full report. Reasons for this included a larger sample size in the full publication and a misrepresentation in the abstract of the number analysed as compared to the number randomized.

In the present study, comparing the information reported about a trial in a conference abstract with that reported in its subsequent full publication, a number of quality issues

were also identified. The method of allocation concealment was not reported in any of the conference abstracts and trials rarely reported the method of blinding in either the conference abstract or the full publication. Intention-to-treat analysis was also poorly described, especially in those trials reported as conference abstracts. The majority of trials were published prior to the introduction of the CONSORT statement in 1996 (Begg et al, 1996), which aims to improve the quality of reports of randomized trials published as journal articles. Several studies (Moher et al, 2001a; Egger et al, 2001b) have shown that the quality of trials published as journal articles has improved following the introduction of the CONSORT statement. In an assessment of 71 randomized trials published in three journals in 1994, the method of allocation concealment was only reported in 39% of trials. This improved to 61% in 1998 when these journals required authors reporting a randomized trial to follow the CONSORT statement (Moher et al, 2001a). Unfortunately, in this current study, the sample of trials published in full, post CONSORT, was too small to detect any significant changes in reporting over time.

The checklist devised and methods used in this study, to compare the content and quality of cancer trials reported as conference abstracts with their subsequent full publication, has since been used in a subsequent study in Australia (Tooher et al, 2004) (Appendix 4). The sample of trials were those identified in the proceedings of three international surgical meetings. Forty-five reports of randomized trials of surgical interventions were identified from the proceedings of the meetings, of which 39 trials had subsequently been published in full. This study of surgical trials found that

although conference abstracts lacked methodological detail and were often interim reports with fewer participants than the full publication, the direction of the results and the conclusions drawn were rarely different - thus adding support to the findings of this study of randomized trials in cancer.

Given the inconsistency between the information reported in the conference abstract and full publication, systematic reviewers and others are still faced with the dilemma as to whether to include information about a trial from a conference abstract in their review, if no full publication is available. Based on the results of this and previous studies (Bhandari et al, 2002; Chokkalingam et al, 1998; Weintraub, 1987) the evidence suggests that information about the participants, the interventions and the primary outcome tend to be stable between the abstract and the full publication. It should therefore be possible to determine reliably if a trial reported as an abstract is eligible to be included in a systematic review providing, of course, that the interventions and outcomes of interest are presented in the abstract.

The evidence also shows that numerical data about the participants included and analysed in the trial can be less reliable between the abstract and full publication, especially if the conference abstract reports early or preliminary results. It is recommended that those carrying out systematic reviews should seek more information from the authors of a trial reported in an abstract. This, in itself, can be problematic and can be a time-consuming part of the systematic review process (Hadhazy et al, 1999). A randomized trial, called "First Contact" is currently being conducted which aims to

identify the most appropriate means of obtaining additional information from the trial investigators (Higgins, 2004). The trial is comparing a standard letter with a more intensive approach involving advance warning, additional information, follow-up and other techniques aimed at improving response rates.

If it is not possible to obtain more information from the trialists, it is suggested that numerical data be included in a systematic review with caution. Those involved in writing the review should make it clear which information relates to a trial reported as a conference abstract. If included in the meta-analysis part of the systematic review, it is recommended that a sensitivity analysis is always performed to assess how the results of the trial reported in the abstract differ from the overall analysis.

In addition, systematic reviewers who are not able to obtain a copy of the full publication of the trial and have only the abstract of the full publication are faced with a similar dilemma to those who only have copy of a conference abstract. The comparison of information reported for a trial in a conference abstract with that which was reported in the abstract of the full publication revealed, surprisingly, that in many instances the abstracts of full publications were more poorly reported than the corresponding conference abstracts. Abstracts of journal articles are generally available in electronic databases such as MEDLINE and EMBASE and, as such, need to accurately report the trial methodology and its results accurately to allow those searching these databases to identify relevant studies more easily (Lefebvre and Clarke, 2001). Well reported abstracts are also important for those conducting systematic

reviews of these trials as reviewers often base their initial assessment of whether a trial is eligible for inclusion in a review on the information reported in an abstract (Alderson et al, 2003). They then use this information to decide whether or not to obtain a copy of the full article.

Even when it is possible to obtain a copy of the full publication, the article may not always be in English; as was the case for the trials which had to be excluded from this study (Figure 6.1). This again emphasises the importance of a well structured abstract as part of a full journal article. The CONSORT statement says that it encourages the use of structured abstracts, which include a series of headings pertaining to the design, conduct and analysis of a trial but this is not a formal requirement (Moher et al, 2001b). Given the potentially stand-alone nature of abstracts of journal articles there is a need for them to be better reported. The CONSORT statement seems the obvious route through which this could be achieved.

CHAPTER 7: DISCUSSION, IMPLICATIONS AND CONCLUSIONS

Introduction

The objectives of this thesis were to assess the actual and potential impact of including trials found in the grey literature in systematic reviews of healthcare interventions. The first chapter of this thesis provided detailed background material on what is grey literature and why it is important to include trials from the grey literature in systematic reviews. Chapter Two described the results of a systematic review of the current knowledge in this area of research. The following chapters described four substantial research projects that were conducted to examine the effects of including grey literature in systematic reviews to build on the current state of the evidence. This final chapter will bring together the findings from each of these research projects, discuss the implications of these findings and present possible ways in which potential problems could be addressed. Possibilities for future research and conclusions will then be presented.

Discussion of main findings

The results of systematic reviews are playing an increasing role in the conduct of evidence-based medicine both in the UK and throughout the rest of the world. The validity of the results of a systematic review is highly dependent on the results of the underlying data and, as such, requires the identification of all relevant studies of a

particular healthcare intervention. Without the complete as possible identification of all relevant studies, the results and conclusions of that systematic review may be biased. If results of biased systematic reviews are then used in the conduct and delivery of health care, interventions will be used without solid evidence as to their effectiveness and possible harms.

Impact of randomized trials reported in the grey literature

The evidence presented in this thesis shows that there may be a systematic difference between the results of randomized trials which are published as full journal articles, and therefore easily accessible to clinicians, systematic reviewers and others, and those trials which are not formally published and therefore difficult to identify. Such difficult to identify sources include trials reported as abstracts in the proceedings of scientific meetings, government and pharmaceutical reports, theses and dissertations, personal correspondence with the trialists and other unpublished sources sometimes classified as "file drawer" data. These difficult to identify sources are termed "grey" literature sources (Auger, 1998).

Trials identified in the published literature (i.e. trials published as full journal articles) tended to be larger and were more likely to show a greater effect of the treatment intervention than trials identified in the grey literature; of which, trials reported as abstracts in the proceedings of scientific meetings, were the principal source (Hopewell et al, 2004c). Excluding trials reported in the grey literature may have particular implications in meta-analyses containing only a small number of trials, as this is where

the impact of not including trials from the grey literature has the greatest potential to introduce bias.

The evidence suggests that the impact of trials found in the grey literature may vary depending on the healthcare condition or intervention under investigation. The findings from this thesis show that it is not possible to predict in advance how large the effect of excluding trials from the grey literature will be, nor is it possible to provide a blanket ruling such as to say that you need to search for grey literature trials in one area of health care but not in another. This means that those carrying out systematic reviews always need to search for trials identified in both the grey as well as the published literature. It is not until such trials have been identified and included in a systematic review that their effect on the overall results and conclusions of a review will become apparent.

These findings are unable to predict how many trials reported in the grey literature, as opposed to the published literature, one might be expected to find for a particular healthcare intervention and how many trials have been missed. It is relatively easy to identify reports of published trials, as they are often indexed in bibliographic databases such as MEDLINE and EMBASE, although it should be recognised that there are still potential problems with this approach (Hopewell et al, 2004a). However, by their nature, trials reported in the grey literature are much more difficult to identify. If the nature and direction of the results were similar for both the published and grey trials, then it would not matter if trials in the grey literature were not identified. However,

there is considerable evidence to show that trials with statistically significant results are much more likely to be published as full journal articles than trials with statistically non-significant or null results (Dickersin, 1997; Cooper et al, 1997; Dickersin and Min, 1993; Dickersin et al, 1992; Easterbrook et al, 1991).

Problems in identifying the evidence

When assessing the findings from a systematic review it is important to look at the search that has been conducted and whether there was an attempt to search for grey as well as published trials. In a Cochrane systematic review, the authors are required to state explicitly the sources that have been searched and when the searches were conducted (Alderson et al, 2003). This then allows the reader independently to assess the validity of the results and its conclusions. However, this level of detail and transparency is not present in all systematic reviews and it can therefore be difficult to assess the reliability of the review findings (Mulrow, 1994).

In the assessment of Cochrane systematic reviews in cancer, it was reassuring that the authors of reviews do appear to search for and include trials from the grey as well as published literature in their reviews. However, in both this assessment and in the review of empirical studies assessing the impact of grey literature in meta-analyses, the percentage of grey trials was still comparatively low compared to the number of published trials included in a systematic review. It was anticipated that the number of trials identified in the grey literature and included in a Cochrane review might have

been higher given the detailed level of searching required. It is possible that despite rigorous searching some trials from the grey literature might not have been identified.

Detecting missing reports of randomized trials

Several statistical methods have been advocated as a way of assessing possible missing studies or publication bias in a systematic review. Perhaps the most commonly known statistical method for detecting publication bias is called the funnel plot. This is a simple scatter plot where the treatment effects estimated from individual trials are plotted on the horizontal axis and some measure of the size of the trial (such as its standard error) is plotted on the vertical axis. If there is no evidence of publication bias then the plot should be shaped like a funnel (Sterne et al, 2000). However, if only studies with significant results are published and included in the meta-analysis then there will be an empty area at the bottom of the funnel.

Another method for detecting missing trials in a meta-analysis is Rosenthal's fail-safe N method, which attempts to estimate the number of possible unpublished trials in a meta-analysis. It does this by estimating the number of unpublished trials required, with a zero treatment effect, to overturn a significant result in a meta-analysis (Rosenthal, 1979). If the number of unpublished trials required is large and therefore, unlikely to exist, then the impact of publication bias is considered to be negligible and thus the results obtained from published trials are judged to be reliable. Other such methods include the trim and fill method, the rank correlation test and the linear regression approach (Sutton et al, 2000). It should, however, be stressed that all of

these approaches have their limitations, as they rely heavily on certain assumptions, which can be difficult to verify in the real world. As such, it has been recommended that such methods should be used with caution and only for the purposes of sensitivity analyses (Song et al, 2000).

Overcoming the problems of publication bias

Ultimately, unless one knows the number of trials that have been carried out for a particular healthcare intervention, irrespective of their final publication status, one will not know how many trials have been missed, despite the rigorous searches conducted. One suggestion that has been posed many times to address the problems of publication bias is the prospective registration of trials at their inception (Tonks, 2002; Chalmers, 2001; Chalmers and Altman, 1999; Horton and Smith, 1999; Simes, 1986). Prospective registration of ongoing trials would enable systematic reviewers and others using the evidence from clinical trials, to identify all the trials that have been conducted which are relevant to a particular healthcare question. In addition, research-funding agencies (public, commercial and charitable), would be able to make their funding decisions in the light of information about trials that are already being conducted. This would, therefore, avoid unnecessary duplication of effort and help to promote collaboration between individual organisations. Finally, if information about ongoing trials were publicly available, patients and clinicians would be better informed about trials in which they could participate which would in turn also help to increase trial recruitment rates.

Several prospective trial registers currently exist, the largest of which is the *meta*Register of Controlled Trials (www.controlled-trials.com). This is an international database containing 27 different trial registers from four continents and containing over 6,000 records of randomized trials in all areas of health care. The registers included in this database are primarily those funded by public funding bodies such as the National Health Service Research and Development Programme in the United Kingdom and the National Institutes of Health in the United States or trials funded by charitable organisations. In the late 1990s two pharmaceutical companies, GlaxoWellcome and Schering Health Care, agreed to publish information about their future clinical trials programmes (Sykes, 1998). Sadly, since GlaxoWellcome's merger to become GlaxoSmithKline details of their trials were no longer made publicly available. However, a recent legal case in the United States, accusing the company of fraudulently withholding information about an antidepressant called Paxil, appears to be instrumental in reversing this policy (http://ctr.glaxowellcome.co.uk/e). An initiative by the Association of the British Pharmaceutical Industry (ABPI) has also seen the development of a clinical trials database, which is a freely available and searchable retrospective register of phase III clinical trials and ongoing phase IV clinical trials in the United Kingdom (www.cmrinteract.com/clintrial/).

One of the most important ways to improve the accessibility of trial data is a simple method of identifying trials throughout their lifecycle (Antes and Dickersin, 2004; Dickersin and Rennie, 2003). The introduction of an International Standard Randomized Controlled Trials Number (ISRCTN) scheme was conceived to address

the confusion experienced by the clinical trials community in which several trials may have the same title, one trial may be reported in several places under different titles, and many trials are never reported. The ISRCTN register assigns a unique number to each randomized trial in order to track the trial unambiguously throughout its lifecycle (http://controlled-trials.com/isrctn/).

It is important to note that trial registration alone does not necessarily mean accurate and complete reporting of trials. There is a need to change the attitudes of journal editors to convince those conducting randomized trials that trials with statistically non-significant results, which are adequately powered, are still equally worthy of publication. The traditional view of medical journal editors has been that they will only publish the results of highly significant studies because these are the articles that would be of most interest to their readers (Cook et al, 1993). This attitude, however, only extenuates the presence of publication bias. One initiative, which aims to break with this tradition, is the *Journal of Negative Results in Biomedicine*. This online journal, published by *BioMed Central* (www.biomedcentral.com), includes papers on all aspects of unexpected, controversial and negative results and conclusions and says that it aims to provide scientists and physicians with responsible and balanced information to support informed clinical decisions (www.jnrbm.com/home/).

Other initiatives are being developed which go beyond the realms of the traditional medical journal. The use of the Internet and electronic journals such as those published by *BioMed Central* (www.biomedcentral.com) allow authors to publish material more

easily. Electronic publishing allows material such as the research protocol and additional information such as data tables to be published that would not normally be included within a printed journal because of the limitations of space. *The Lancet* has also developed an initiative to publish the protocol for a trial on its web site before the trial has been completed and its results published in full (www.thelancet.com). In addition, some professional organisations, such as the American Society of Clinical Oncology, are making available and searchable the abstracts from their scientific meetings on their organisation's web site. This then enables those who did not attend the meeting or who do not have access to a copy of the conference proceedings, to obtain information about trials presented at the meeting.

Currently, there is no legal requirement to register and make publicly available the full results of clinical trials. Without this mandatory requirement large numbers of trials will continue to be susceptible to selective and poor reporting and thus extenuating the problems of publication bias. It has been argued that to conduct clinical research and not make publicly available the results of that trial is a serious scientific misconduct (Chalmers, 1990). Those individuals taking part in a trial do so on the understanding that there was uncertainty in the current state of the evidence, otherwise known as clinical equipoise, thereby taking part in the trial would help guide future practice and benefit other individuals, such as themselves, in the future. Initiatives, which will encourage people to publish the results of clinical research more easily and therefore make them more assessible to their users can only be a good thing and should be encouraged.

Quality of randomized trials reported in the grey literature

Irrespective of whether a trial is identified in the grey or published literature, it is important to assess the quality of its methods, as trials of poorer methodological quality lead to exaggerations in the effect of the treatment intervention (Egger et al, 2003; Kjaergard et al; 2001; Moher et al, 1998; Schulz et al, 1995). It is recommended that those carrying out a systematic review routinely conduct a sensitivity analysis as part of the meta-analysis to assess the impact of the quality of individual trials on the overall results of a review (Moher et al, 1999b).

It had previously been suggested that trials identified in the grey literature might be of poorer methodological quality than those published as full journal articles (Egger et al, 2003). However, in the assessment of trials reported in the grey literature and included in Cochrane reviews presented in this thesis, there appeared to be no significant difference in the reporting of methodological quality for both published and grey trials. In this assessment, the methods of generation and concealment of the allocation sequence and the method of blinding were frequently unclear or not reported for a large proportion of published and grey trials. In comparison, published trials where additional information had been obtained directly from the trialists, tended to be categorised as of higher methodological quality. This illustrates the importance of contacting trialists for more information where this is either incomplete or missing.

Poor reporting of methodological quality was also apparent in the assessment of trials reported in conference abstracts and their subsequent full publications. While details of blinding and the randomization process were poorly reported in the conference abstracts, it was not anticipated that this should also be poorly reported in the full publications. One of the obvious limitations of an abstract is the space in which to describe key aspects of the trial. A trial published as a full journal article does not face the same space limitations as a trial reported in an abstract and therefore the authors should be able to provide a detailed description of the trial methodology in the full publication.

Improving the quality of randomized trials

The Consolidated Standards of Reporting Trials, more commonly known as the CONSORT statement, is an important tool to improve the quality of reports of randomized trials. The CONSORT statement was originally published in 1996 (Begg et al, 1996) and revised in 2001 (Moher et al, 2001b) and aims to help improve the quality of reports of randomized trials published in medical journals. It thereby allows readers to understand better the design, conduct, analysis and interpretation of the trial results. CONSORT comprises of a checklist and flow diagram. The checklist includes items, based on evidence, that need to be addressed in the report and the flow diagram provides readers with a clear picture of the progress of all participants in the trial, from the time they are randomized until the end of their involvement. The intention is to make the process of the trial clearer, flawed or not, so that users of the data can evaluate, more appropriately, the validity of the trial results. The CONSORT statement

is now available in several languages and has been endorsed by prominent medical journals such as *The Lancet*, *Annals of Internal Medicine* and *JAMA*. As of July 2004, 163 journals have endorsed the use of the CONSORT statement for reporting the results of randomized trials in their journals (www.consort-statement.org).

Providing authors with specific instructions does appear to improve the quality of the reporting of randomized trials. In a before and after study, a sample of randomized trials published in 1996 was compared with a similar sample published in 1998 (Moher et al, 2001a). Three journals were sampled, the *BMJ*, *JAMA* and *The Lancet* all of which had endorsed the CONSORT statement from its inception in 1996. One journal, the *New England Journal of Medicine*, was sampled, which had not endorsed the CONSORT statement and thus acted as a control. As one might expect, there was an increase in the number of items from the CONSORT checklist that were included in the reports from the endorsing journals. This was not observed in the reports from the non-endorsing journal. Trial quality was assessed using the Jadad scale, which is a five-point quality assessment instrument. This showed a significant improvement in the quality assessment score for the trials reported in the endorsing journals but not for the non-endorsing journal.

Differences between poor reporting and poor quality

It is important to recognise the distinction between poor reporting and poor quality. There is some evidence to suggest that a poorly reported trial does not necessarily mean that it is of poor methodological quality (Soares et al, 2004; Huwiler-Muntener et

al, 2002; Liberati et al, 1986). One such study compared the methodological quality of published oncology trials with that described in the original research protocol. Soares and colleagues (2004) found that adequate allocation concealment was achieved in all trials but was only reported in 42% of the published articles, sample size calculations were performed in 76% of trials but only reported in 16% of published articles and, intention-to-treat analysis was done in 83% of trials but reported in only 69% of the published articles. Those conducting systematic reviews of randomized trials rarely have access to the original protocol for a trial. However, these findings again help to highlight the importance of contacting trialists for additional information when key aspects of the trial are either not reported or are unclear, irrespective of a trial's publication status.

Reporting of randomized trials in the grey literature

A large number of trials are presented at scientific meetings and reported as abstracts in the proceedings of the meeting. Only just over half of these trials are subsequently published in full, thus making conference abstracts a principal source of grey literature. Searching by hand the proceedings of the American Society of Clinical Oncology annual conference and the Thoracic Society of Australia and New Zealand annual conference identified a large number of trials which had not subsequently been published as full journal articles and which potentially would be eligible for inclusion in systematic reviews. However, one of the potential problems of trying to identify and include trials reported in conference abstracts in a systematic review is the limited amount of information reported about that trial. While some aspects of a trial appear

well reported in conference abstracts, such as details of the trial objectives, participants and interventions, other aspects are much more poorly reported. Particular aspects that were very poorly reported included details of the methods of the trial and information about the participants randomized and then included in the analysis. In addition, an increasing number of abstracts appear to be reporting interim trial analysis or, indeed, state that the results of the trial will be presented at the meeting. When the information reported for a trial in a conference abstract was compared with that reported in its subsequent full publication, this reinforced the preliminary nature of some of the trials presented in conference abstracts.

One of the potential reasons for the increasing number of trials reporting interim analyses or, indeed, no results in the conference abstract, is the pressure to present the results of a trial sooner and potentially before the true effect of the intervention is known (Lexchin et al, 2003). Pressure to publish means that the results of trials with more striking findings become available sooner than those with less striking findings (Hopewell et al, 2004b). Therefore, if the time taken to publish trials with significant results is shorter than it is for those with non-significant or null results, new interventions will be accepted as effective in the absence of evidence to the contrary, even though that evidence may already have been gathered (Jadad, 1998). This may then introduce bias into a systematic review if a trial's inclusion is related to the timing of the availability of its data. It could, therefore, be argued that a trial should only be accepted for presentation at a scientific meeting, if full details of the trial, its methodology and its results are presented in the initial abstract.

There are, however, problems with this approach as there may be occasions when it is inappropriate or indeed impossible for trialists to report the results of a trial in the initial conference abstract, which may well be submitted five or six months before it is presented at the meeting. An alternative approach would be to allow trialists to submit an updated version of their abstract, which contains result data, one or two months before the meeting. If this is not possible, there needs to be a mechanism for making this information publicly available directly after the meeting and without waiting for the trial to be published in full if it is indeed published at all. One suggestion, which draws on the ideas discussed earlier in this chapter, is to make greater use of the Internet and electronic publishing. Each conference abstract could include a link to the trial's web site where the trial results or, at a minimum, the slides presented at the meeting, should be posted. In addition, if all trials were allocated an ISRCTN number, which was also reported in the conference abstract, this would allow those assessing abstracts for inclusion in a systematic review to track the results of a trial, if they are not presented in the initial conference abstract.

Improving the reporting of randomized trials

Despite initiatives such as prospective trial registration and the use of the Internet, it is recognised that not all trials will be published in full and might only ever be available in the proceedings of the conference at which the trial was presented. The evidence presented in this thesis has highlighted the importance of contacting the trialists for more information about a trial, where this is either incomplete or missing. Despite this,

it is argued that an abstract should be able to serve as a lasting record of that trial, irrespective of whether it is published in more detail elsewhere. Without basic information about a trial, it is not possible to assess the validity of its results and the importance of its conclusions.

Similar to the CONSORT statement for reporting randomized trials in journal articles, it is proposed that a modified version of CONSORT (such as a mini-CONSORT or another such pseudonym) should be developed for reporting randomized trials in the proceedings of scientific meetings. This tool could also set out to define what is an acceptable word limit for reporting the results of a randomized trial in a conference abstract. In the sample of abstracts assessed in this thesis, the typical word limit was approximately 250 to 300 words.

A statement for the reporting of randomized trials in the proceedings of scientific meetings would serve several purposes; firstly it would allow those assessing abstracts for presentation at the meeting to have a better understanding of how the trial was conducted and therefore the quality of its results. This in turn should lead to a higher standard of trials being presented at the meeting and thus help raise the professional credibility of the scientific conference. The second benefit of a tool such as the CONSORT statement would be to allow users of the abstracts, such as those conducting systematic reviews of this evidence, to do so more effectively and thereby assess their validity more closely.

Developing a tool to improve the quality of randomized trials reported in abstracts

The checklist developed as part of this thesis, which was adapted from existing quality criteria such as the CONSORT statement (Moher et al, 2001b) and the structured reporting of journal abstracts and short reports (Deeks and Altman, 1998; Haynes et al, 1996; Haynes et al, 1990; Mulrow et al, 1988), provides an ideal precursor to the development of a mini-CONSORT statement. If such a statement is to be recognised and implemented internationally, it needs to be widely endorsed by the scientific community. It is therefore advocated that a similar approach to that involved in the development of the CONSORT statement should be adopted. Examples of other such tools which have been developed to improve the quality of reporting include: the QUOROM statement (Quality Of Reporting Of Meta-analyses) for the reporting of systematic reviews and meta-analyses (Moher et al, 1999a); the MOOSE initiative (Meta-analysis of Observational Studies in Epidemiology) for reporting of meta-analyses of observational studies (Stroup et al, 2000); the STARD initiative (Towards Complete and Accurate Reporting of Studies on Diagnostic Accuracy) for the reporting of studies of diagnostic accuracy (Bossuyt et al, 2004); and the ASSERT statement (A Standard for the Scientific and Ethical Review of Trials) for the review and monitoring of randomized trials by research ethics committees (www.assert-statement.org/).

These different initiatives were developed by international groups of clinical trialists, epidemiologists, statisticians and biomedical editors and are supported by a growing number of medical and healthcare journals. Each of the initiatives just described was

set up with their own steering committee and the tools were developed through a series of intensive meetings and workshops. In some instances, the Delphi process was used to provide and collate an extensive list of potential items, which could then be scored in order of scientific importance and then used in the final quality tool. It is proposed that a similar international group would be required to develop a modified version of the CONSORT statement, which would be specifically aimed at improving the quality of reporting of randomized trials in the proceedings of scientific meetings. Such a group would also require the involvement and support of scientific bodies involved in the organisation of large international clinical trial meetings.

It would be a requirement of these large scientific meetings for trialists to submit abstracts reporting the results of randomized trials using the agreed new reporting criteria. If the trialists did not comply with this requirement, then their work might not be accepted for publication or presentation at the meeting irrespective of how interesting the results of that trial might be; as would be the case for a trialist submitting a manuscript to a journal that endorsed the CONSORT statement. The obvious down side to this approach is that other research, such as observational studies, are also submitted and presented at scientific meetings and these would not be covered by this guidance and may also be of poorer methodological quality (Deeks et al, 2003). An alternative approach, which might be considered, is to develop more global reporting criteria for conference abstracts, which would also incorporate information for other types of study design.

Implications for systematic reviewers

The findings from this thesis have important implications for systematic reviewers, those who conduct randomized trials and others including the users of evidence from trials and reviews. Implications for those involved in the conduct of systematic reviews of randomized trials include:

- Those carrying out systematic reviews must ensure they search for reports of randomized trials identified in both published and grey literature sources. These sources include searching electronic bibliographic databases, handsearching key journals, searching conference proceedings, contacting researchers, searching research registers and searching the Internet.

- Whatever sources are searched, it is important that systematic reviewers are explicit and state in the review which sources they have searched and when the search was conducted. This then allows the reader to assess the validity of the conclusions of the review based on the evidence that has been identified.

- One of the obvious limitations of trials reported in the grey literature, specifically those reported as abstracts in the proceedings of scientific meetings, is the limited amount of space in which to report the results of a randomized trial. It is important that systematic reviewers assessing these reports determine the difference between a poor quality trial and poor reporting of a trial's methods and its results.

- Those conducting systematic reviews should ensure that they assess the quality of individual reports of randomized trials identified for inclusion in a systematic review, irrespective of whether they are identified in the published or grey literature.

- Where information about a trial's methodological quality and or its results is either incomplete or missing, it is important that systematic reviewers contact the trialists for additional information. This should apply to all trials irrespective of their publication status.

- Finally, it is recommended that those conducting systematic reviews routinely conduct a sensitivity analysis as part of any meta-analysis to assess the impact of methodological quality and source of publication of individual trials on the overall results of the systematic review.

Implications for trialists

Implications for those involved in the conduct and delivery of randomized trials include:

- There is a need for trialists to improve the quality of reporting of randomized trials submitted as abstracts and published in the proceedings of scientific meetings. A trial reported in a conference abstract should contain enough information about the

trial methodology and its results to allow readers independently to assess the validity of the trial's conclusions.

- There is also a need to increase the accessibility of randomized trials reported in conference proceedings. Posting the proceedings of the meeting on the Internet would allow those who did not attend the meeting or who do not have access to the meeting abstract book, to identify reports of these trials more easily.

- Trialists and journal editors should be strongly encouraged to publish the results of randomized trials, which have been conducted appropriately, irrespective of the statistical significance of their results.

- Where reports of randomized trials are published in medical journals, there is a need to improve the reporting of these trials, in particular the information reported in the abstract of a journal article. This thesis supports the use of the CONSORT statement, however, it also recommends that CONSORT, trialists and journal editors support the use of structured journal abstracts to help improve the quality of reporting of randomized trials.

Implications for users of trial evidence

These findings also have implications for the users of systematic reviews and the randomized trials they incorporate. These users include patients and clinicians who want to interpret the findings from individual systematic reviews to influence directly

their treatment decisions. It may also include guideline developers and policy-makers involved in evidence-based medicine decision-making. Possible implications include:

- Those assessing the findings from a systematic review need to consider whether the conclusions of the review are valid, based on the searches that were conducted. It is important to assess whether the systematic review attempted to analyse all available evidence for a particular healthcare intervention and not just evidence from published sources.

- Information about ongoing trials and trials identified in the grey literature need to be publicly available so that patients and clinicians are better informed about trials in which they can participate. This would also help increase recruitment rates and reduce duplication of effort and the effects of publication bias.

- It is argued that prospective registration of all randomized trials at their inception, and the public availability of results of trials at their completion, should be a formal requirement by all funding bodies and medical ethics committees. Not to do so should be considered as scientific misconduct and fraudulent.

Implications for future research

From this thesis several questions remain that may still warrant further investigation. The reports of randomized trials assessed in this thesis were from specific areas of health care. Where it was available, evidence has been systematically gathered and

presented to show how these findings are relevant to and supported by evidence from other areas of health care. Still, it is possible that the findings from this thesis may not be transferable to all areas of health care. In particular, it should be recognised that the sample of trials gathered from the proceedings of scientific meetings are largely those reported at a large international cancer meeting. It is therefore possible that the evidence presented, comparing the quality and reporting of trials in conference abstracts, would be different in other areas of health care. Indeed, the picture could be worse in smaller less prestigious areas of medicine, where the science is not developing as rapidly. Further research might therefore be required to investigate specific aspects of the quality and reporting of trials identified in the grey literature in areas of health care other than those assessed in this thesis.

Even when a trial is subsequently published in full, there is evidence of selective reporting of the outcomes of the trial when compared to what was reported in the original trial protocol (Chan et al, 2004). The study by Chan and colleagues found that trial outcomes with statistically significant results were more likely to be reported in full than trial outcomes with non-significant results. If the reporting of trial outcomes are incomplete or inconsistent with those reported in the trial protocol then the published article and the reviews that incorporate them may be unreliable and overestimate the benefits of a treatment intervention.

The opportunity for selective reporting of outcomes is potentially even greater in the reporting of trials in the proceedings of scientific meetings. Given the limited amount

of space available in an abstract, in which to describe the results of a randomized trial, the incentive to report only the significant "eye catching" outcomes and results cannot be underestimated. There is no strong evidence to support this claim but some indirect evidence can be drawn from this thesis when comparing the information reported for a trial in a conference abstract with that reported in its subsequent full publication. In a small number of cases, the conclusions of a trial were significantly different in the full publication compared to the initial conference abstract. Further research is needed to explore this possibility of selective reporting. Such research might assess the life history of a trial, following it from its inception at the initial trial protocol stage, then at the abstract presentation stage, to its final full publication. The study would look at which outcomes are reported at each of these stages of the life of the trial and the significance of their results.

Conclusions

Several important conclusions can be drawn from this thesis. Firstly, published trials tend to be larger and are more likely to show larger effects of a treatment intervention than those trials identified in the grey literature. The extent of this difference between published and grey trials varies depending on the area of health care, the treatment intervention under investigation and the time period in which the trial was done. Those carrying out systematic reviews therefore need to search for and include trials from the published and grey literature. Failure to do so may result in the results and conclusions of that review being biased and unreliable.

Secondly, previous evidence had suggested that trials found in the grey literature were of poorer methodological quality. This thesis found very few important differences in the methodological quality of published and grey trials and that this information is often not reported or unclear in both of these sources. However, this thesis does confirm that trials of poorer methodological quality are more likely to show larger effects of a treatment intervention than trials reporting adequate methods.

This thesis identified that some aspects of trials reported in the grey literature, specifically those reported as abstracts in the proceedings of scientific meetings, lack important information about trial results and the participants included and analysed in the trial. This appears to be getting worse in recent years as the pressure to present the results of a trial sooner increases and the tendency for abstracts to be "promissory notes" of what will be presented at the meeting itself increases. The quality of conference abstracts containing reports of randomized trials needs to be improved. It is proposed that a reporting standard, similar to the CONSORT statement for reporting randomized trials as journal articles, is developed to address this problem of poor and incomplete reporting of trials in the proceedings of scientific meetings.

It is important that systematic reviewers can distinguish between a poor quality trial and poor reporting of a trial's methods and its results. Systematic reviewers need to document in a review what sources have been searched and when, so that readers can assess its validity. Where information about a trial is either not reported or unclear, irrespective of whether it is published in full or not, it is important to obtain additional

information from the trialists. It is recommended that a sensitivity analysis be routinely conducted as part of a meta-analysis to assess the impact of the source of individual trials and their quality on the overall results of a systematic review.

Finally, it is proposed that prospective registration and mandatory reporting of all randomized trials remains the most appropriate mechanism to minimise or even prevent the threat of publication bias and the problems of poor and incomplete reporting identified in this thesis. Without this mandatory requirement, trials reported only in the grey literature remain a threat to the validity and reliability of systematic reviews, evidence-based health care and the clinical decisions made based on the findings from these systematic reviews.

REFERENCES

Alberani, V., de Castro Pietrangeli, P. and Mazza, A.M. (1990) The use of grey literature in health sciences, a preliminary survey. *Bulletin of the Medical Library Association*, 78(4), 358-363.

Alderson P, Green S, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003]. In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd. (Citation as instructed).

Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P.C. and Lang, T. (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663-694.

Anon. (2004) EMBASE [Release Notes]. *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley & Sons, Ltd.

Antes, G. and Dickersin, K. (2004) Trial registration to prevent duplicate publication. *JAMA*, 291(20), 2432.

Antes, G. and Oxman, A.D. (2001) The Cochrane Collaboration in the 20th century. *In:* M. Egger, G. Davey Smith and D.G. Altman, editors. *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd edition. London, BMJ Books. p. 447-458.

Auger, C.P. (1998) *Information sources in grey literature (guides to information sources)*. 4th edition. London, Bowker Saur.

Bardy, A.H. (1998) Bias in reporting clinical trials. *British Journal of Clinical Pharmacology*, 46(2), 147-150.

Bhandari, M., Devereaux, P.J., Guyatt, G.H., Cook, D.J., Swiotkonwski, M.F., Sprague, S. and Schemitsch, E.H. (2002) An observational study of orthopaedic abstracts and subsequent full-text publications. *Journal of Bone and Joint Surgery American*, 84-A(4), 615-621.

Bhandari, M., Guyatt, G.H., Tong, D., Adili, A. and Shaughnessy, S.G. (2000) Reamed versus nonreamed intramedullary nailing of lower extremity long bone fractures: a systematic overview and meta-analysis. *Journal of Orthopaedic Trauma*, 14(1), 2-9.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D. and Stroup, D.F. (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, 276(8), 637-639.

Boruch, R.F., McSweeney, A.J. and Soderstrom, E.J. (1978) Randomized field experiments for program planning, development, and evaluation: an illustrative bibliography. *Evaluation Quarterly*, 2(4), 655-695.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D. and de Vet, H.C. (2004) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Family Practice*, 21(1), 4-10.

British Library. Document Supply Centre. *Catalogues and collections - grey literature* [online]. Boston Spa, Wetherby, British Library Document Supply Centre. Available from: www.bl.uk/services/document/greylit.html [Accessed 5 December 2003].

Burdett, S., Stewart, L.A. and Tierney, J.F. (2003) Publication bias and meta-analyses: a practical example. *International Journal of Technology Assessment in Health Care*, 19(1), 129-134.

Callaham, M.L., Wears, R.L., Weber, E.J., Barton, C. and Young, G. (1998) Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA*, 280(3), 254-257.

Chalmers, I. (1990) Underreporting research is scientific misconduct. *JAMA*, 263(10), 1405-1408.

Chalmers, I. (2001) Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: M. Egger, G. Davey Smith and D.G. Altman, editors. *Systematic Reviews in Heath Care: Meta-Analysis in Context*. 2nd edition. London, BMJ Books. p.429-443.

Chalmers, I. and Altman, D.G. (1999) How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet*, 353(9151), 490-493.

Chan, A.W., Hrobjartsson, A., Haahr, M.T., Gøtzsche, P.C. and Altman, D.G. (2004) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*, 291(20), 2457-2465.

Chlebowski, R.T. and Lillington, L.M. (1994) A decade of breast cancer clinical investigation: results as reported in the Program/Proceedings of the American Society of Clinical Oncology. *Journal of Clinical Oncology*, 12(9), 1789-1795.

Cho, M.K. and Bero, L.A. (1996) The quality of drug studies published in symposium proceedings. *Annals of Internal Medicine*, 124(5), 485-489.

Chokkalingam, A., Scherer, R. and Dickersin, K. (1998) Agreement of data in abstracts compared to full publications. *Controlled Clinical Trials*, 19(Suppl 1), 61-62S.

Ciesla, M.C. and Wojcik E.M. (2001) Outcomes of cytopathology studies presented at national pathology meetings. *Diagnostic Cytopathology*, 25(4), 265-269.

Clarke, M. and Clarke, T. (2000) A study of the references used in Cochrane protocols and reviews. Three bibles, three dictionaries, and nearly 25,000 other things. *International Journal of Technology Assessment in Health Care*, 16(3), 907-909.

Clarke, M. and Greaves, L. (1996) Searching meeting abstracts: a comparison between an experienced handsearcher and an indexed, electronic database. *In: Fourth International Cochrane Colloquium, 1996 Oct 20-24, Adelaide, Australia.* p. A51.

Cochrane, A. (1972) *Effectiveness and efficiency.* London, The Royal Society of Medicine Press Limited.

Cook, A.M., Finlay, I.G., Edwards, A.G., Hood, K., Higginson, I.J., Goodwin, D.M., Normand, C.E. and Douglas, H.R. (2001) Efficiency of searching the grey literature in palliative care. *Journal of Pain and Symptom Management*, 22(3), 797-801.

Cook, D.J., Guyatt, G.H., Ryan, G., Clifton, J., Buckingham, L., Willan, A., McIlroy, W. and Oxman, A.D. (1993) Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA*, 269(21), 2749-2753.

Cooper, H., DeNeve, K. and Charlton, K. (1997) Finding the missing science: the fate of studies submitted for review by a human subjects committee. *Psychological Methods*, 2(4), 447-452.

Corry, A.M. (1990) A survey of the publication history of randomly selected IADR/AADR abstracts presented in 1983 and 1984. *Journal of Dental Research*, 69(8), 1453-1455.

Crowley, P., Chalmers, I. and Keirse, M.J. (1990) The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. *British Journal of Obstetrics and Gynaecology*, 97(1), 11-25.

Davies, M.W., Dunster, K.R., East, C.E. and Lingwood, B.E. (2002) Fate of abstracts published in the proceedings of the first annual Perinatal Society of Australia and New Zealand Congress in 1997. *Journal of Paediatrics and Child Health*, 38(5), 501-506.

De Smidt, G.A. and Gorey, K.M. (1997) Unpublished social work research, systematic replication of a recent meta-analysis of published intervention effectiveness research. *Social Work Research*, 21(1), 58-62.

Deeks, J.J. and Altman, D.G. (1998) Inadequate reporting of controlled trials as short reports. *Lancet*, 352(9144), 1908.

Deeks, J.J., Dinnes, J., D'Amico, R., Sowden, A.J., Sakarovitch, C., Song, F., Petticrew, M. and Altman, D.G. (2003) Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7(27), 1-173.

Detsky, A.S., Baker, J.P., O'Rourke, K. and Goel, V. (1987) Perioperative parenteral nutrition: a meta-analysis. *Annals of Internal Medicine*, 107(2), 195-203.

Dickersin, K. (1997) How important is publication bias? A synthesis of available data. *AIDS Education and Prevention*, 9(1 Suppl), 15-21.

Dickersin, K., Manheimer, E., Wieland, S., Robinson, K.A., Lefebvre, C. and McDonald, S. (2002) Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Evaluation and the Health Professions*, 25(1), 38-64.

Dickersin, K. and Min, Y.I. (1993) NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials* [Serial Online], Doc 50, 53 paragraphs.

Dickersin, K., Min, Y.I. and Meinert, C.L. (1992) Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA*, 267(3), 374-378.

Dickersin, K. and Rennie, D. (2003) Registering clinical trials. *JAMA*, 290(4), 516-523.

Dickersin, K., Scherer, R. and Lefebvre, C. (1994) Identifying relevant studies for systematic reviews. *BMJ*, 309(6964), 1286-1291.

Donaldson, I.J. and Cresswell, P.A. (1996) Dissemination of the work of public health medicine trainees in peer-reviewed publications: an unfulfilled potential. *Public Health*, 110(1), 61-63.

Dupuy, A., Khosrotehrani, K., Lebbe, C., Rybojad, M., and Morel, P. Quality of abstracts in 3 clinical dermatology journals. *Archives of Dermatology*, 2003, 139(5), 589-593.

Easterbrook, P.J., Berlin, J.A., Gopalan, R. and Matthews, D.R. (1991) Publication bias in clinical research. *Lancet*, 337(8746), 867-872.

Eastwood, S., Derish, P., Leash, E. and Ordway, S. (1996) Ethical issues in biomedical research: perceptions and practices of postdoctoral research fellows responding to a survey. *Science and Engineering Ethics*, 2(1), 89-114.

Egger, M., Davey Smith, G., and O'Rourke, K. (2001a) Rationale, potentials, and promise of systematic reviews. *In:* M. Egger, G. Davey Smith and D.G. Altman, editors. *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd edition. London, BMJ Books. p. 3-19.

Egger, M., Davey Smith, G. and Phillips, A.N. (1997) Meta-analysis: principles and procedures. *BMJ*, 315(7121), 1533-1537.

Egger, M., Jüni, P. and Bartlett, C. and the CONSORT Group (Consolidated Standards of Reporting of Trials) (2001b) Value of flow diagrams in reports of randomized controlled trials. *JAMA*, 285(15), 1996-1999.

Egger, M., Jüni, P., Bartlett, C., Holenstein, F. and Sterne, J. (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment*, 7(1), 1-76.

Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C. and Antes, G. (1997) Language bias in randomised controlled trials published in English and German. *Lancet*, 350(9074), 326-329.

Eysenbach, G., Tuische, J. and Diepgen, T.L. (2001) Evaluation of the usefulness of Internet searches to identify unpublished clinical trials for systematic reviews. *Medical Informatics and the Internet in Medicine*, 26(3), 203-218.

Eysenck, H.J. (1994) Meta-analysis and its problems. *BMJ*, 309(6957), 789-792.

Fergusson, D., Laupacis, A., Salmi, L.R., McAlister, F.A. and Huet, C. (2000) What should be included in meta-analyses? An exploration of methodological issues using the ISPOT meta-analyses. *International Journal of Technology Assessment in Health Care*, 16(4), 1109-1119.

Froom, P. and Froom, J. (1993) Deficiencies in structured medical abstracts. *Journal of Clinical Epidemiology*, 46(7), 591-594.

Glass, G.V. (1976) Primary, secondary, and meta-analysis of research. *Educational Research*, 5, 3-8.

Glass, G.V., McGaw, B. and Lee Smith, M. (1981) Meta-analysis in social research. London, Sage Publications.

Gotzsche, P.C. (1987) Reference bias in reports of drug trials. *British Medical Journal* (Clinical Research Edition), 295(6599), 654-656.

Grayson, L. and Gomersall, A. *A difficult business: finding the evidence for social science reviews* (unpublished draft sent to author by A. Gomersall).

Hadhazy, V., Ezzo, J. and Berman, B. (1999) How valuable is effort to contact authors to obtain missing data in systematic reviews. *In: Seventh International Cochrane Colloquium, 1999 Oct 5-9, Rome, Italy*. p. 33.

Harbourt, A.M., Knecht, L.S. and Humphreys, B.L. (1995) Structured abstracts in MEDLINE, 1989-1991. *Bulletin of the Medical Librarys Association*, 83(2), 190-195.

Haynes, R.B., Mulrow, C.D., Huth, E.J., Altman, D.G. and Gardner, M.J. (1990) More informative abstracts revisited. *Annals of Internal Medicine*, 113(1), 69-76.

Haynes, R.B., Mulrow, C.D., Huth, E.J., Altman, D.G. and Gardner, M.J. (1996) More informative abstracts revisited. *Cleft Palate and Craniofacial Journal*, 33(1), 1-9.

Hetherington, J., Dickersin, K., Chalmers, I. and Meinert, C.L. (1989) Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics*, 84(2), 374-380.

Higgins, J. (2004) *First contact with investigators regarding information for a systematic review: a randomised trial.* Available from: www.mrc-bsu.cam.ac.uk/firstcontact/downloads.html [Accessed 22 June 2004]

Holmes, L., Lusher, A. and Chalmers, I. (2001) *Citation of Cochrane reviews in national and international guidelines and policies* (unpublished draft sent to author by L. Holmes).

Hopewell, S. (2003) Assessing the impact of abstracts from the Thoracic Society of Australia and New Zealand in Cochrane reviews. *Respirology*, 8(4), 509-512.

Hopewell, S. (2004) Trials reported as abstracts and full publications: how do they compare? *Clinical Trials*, 1 (Suppl 1), 216-217.

Hopewell, S. and Clarke, M. (2001) Methodologists and their methods: do methodologists write up their conference presentations or is it just 15 minutes of fame? *International Journal of Technology Assessment in Health Care*, 17(4), 601-603.

Hopewell, S. and Clarke, M. (2003) Quality of trials reported as conference abstracts: how well are they reported? *In: 11th International Cochrane Colloquium, 2003 Oct 26-31, Barcelona, Spain.* p. 38.

Hopewell, S., Clarke, M., Lefebvre, C. and Scherer, R. (2004a) Handsearching versus electronic searching to identify reports of randomized trials (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley & Sons, Ltd.

Hopewell, S., Clarke, M., Stewart, L. and Tierney, L. (2004b) Time to publication for results of clinical trials (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley & Sons, Ltd.

Hopewell, S. and Eisinga, A. (2003). Including trials identified in the grey literature in your review [workshop]. *In: Ninth Annual Meeting for UK Contributors to The Cochrane Collaboration, 2003 Mar 17-18, Warwick, UK.*

Hopewell, S. and Eisinga, A. (2004) Grey literature: issues in searching and using [workshop]. *In: 10th Annual Meeting for the UK Contributors to The Cochrane Collaboration, 2004 Mar 25-26, Edinburgh, UK.*

Hopewell, S., MacKinnon, L. and Clarke, M. (2003) Conference abstracts that contain "possible" randomized trials - are they worth finding? *In: 11th International Cochrane Colloquium, 2003 Oct 26-31, Barcelona, Spain.* p. 34.

Hopewell, S. and McDonald, S. (2003) Full publication of trials initially reported as abstracts in the *Australian and New Zealand Journal of Medicine* 1980-2000. *Internal Medicine Journal*, 33(4), 192-194.

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2002a) Grey literature in meta-analyses of randomized trials - a systematic review. *In: Fourth Symposium on Systematic Reviews: Pushing the Boundaries, 2002 Jul 2-4, Oxford, UK.* p. 44.

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2002b) Grey literature in meta-analyses of randomized trials - a systematic review. *In: 10th International Cochrane Colloquium, 2002 Jul 31 - Aug 3, Stavanger, Norway.* p. 29.

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2004c) Grey literature in meta-analyses of randomized trials of health care interventions (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley & Sons, Ltd.

Horn, J. and Limburg, M. (2002) Calcium antagonists for acute ischemic stroke (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2002. Oxford, Update Software.

Horton, R. and Smith, R. (1999) Time to register randomised trials. *Lancet*, 354(9185), 1138-1139.

Huwiler-Muntener, K., Jüni, P., Junker, C. and Egger, M. (2002) Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*, 287(21), 2801-2804.

Ioannidis, J.P. (1998) Effect of the statistical significance of results on the time to completion and publication of randomised efficacy trials. *JAMA*, 279(4), 281-286.

Jadad, A.R. and Rennie, D. (1998) The randomized controlled trial gets a middle-aged check up. *JAMA*, 279(4), 319-320.

Jefferson, T.O., Alderson, P., Davidoff, F. and Wager, E. (2004) Editorial peer-review for improving the quality of reports of biomedical studies (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley and Sons, Ltd.

Jeng, G.T., Scott, J.R. and Burmeister, L.F. (1995) A comparison of meta-analytic results using literature vs individual patient data. Paternal cell immunization for recurrent miscarriage. *JAMA*, 274(10), 830-836.

Jüni, P., Holenstein, F., Sterne, J., Bartlett, C. and Egger, M. (2002) Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology*, 31(1), 115-123.

Jüni, P., Tallon, D. and Egger, M. (2000) Assessment of the quality of controlled trials in meta-analysis published in leading journals. *In: 3rd Symposium of Systematic Reviews: Beyond the Basics, 2000 Jul 3-5, Oxford, UK.* p.19.

Jüni, P., Witschi, A., Bloch, R. and Egger, M. (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*, 282(11), 1054-1060.

Kjaergard, L.L., Villumsen, J. and Gluud, C. (2001) Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine,* 135(11), 982-989.

Klassen, T.P., Wiebe, N., Russell, K., Stevens, K., Hartling, L., Craig, W.R. and Moher, D. (2002) Abstracts of randomized controlled trials presented at the Society for Pediatric Research Meeting: an example of publication bias. *Archives of Pediatrics and Adolescent Medicine,* 156(5), 474-479.

Lefebvre, C. and Clarke, M. (2001) Identifying randomised trials. In: M. Egger, G. Davey Smith and D.G. Altman, editors. *Systematic Reviews in Heath Care: Meta-Analysis in Context*. 2nd edition. London, BMJ Books. p.69-86.

Lexchin, J., Bero, L.A., Djulbegovic, B. and Clark, O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*, 326(7400), 1167-1170.

Liberati, A., Himel, H.N. and Chalmers, T.C. (1986) A quality assessment of randomized control trials of primary treatment of breast cancer. *Journal of Clinical Oncology*, 4(6), 942-951.

Lind, J. (1753) *A Treatise on the Scurvy. In three parts.* Edinburgh, A Kincaid and A Donaldson.

Lipsey, M.W. and Wilson, D.B. (1993) The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American Psychologist*, 48(12), 1181-1209.

MacLean, C.H., Morton, S.C., Ofman, J.J., Roth, E.A. and Shekelle, P.G. (2003) How useful are unpublished data from the Food and Drug Administration in meta-analysis? *Journal of Clinical Epidemiology*, 56(1), 44-51.

Mallett, S. and Clarke, M. (2002) The typical Cochrane review. How many trials? How many participants? *International Journal of Technology Assessment in Health Care*, 18(4), 820-823.

Mallett, S., Hopewell, S. and Clarke, M. (2002) The use of grey literature in the first 1000 Cochrane reviews. *In: Fourth Symposium on Systematic Reviews: Pushing the Boundaries, 2002 Jul 2-4, Oxford.* p.43.

Man-Son-Hing, M., Wells, G. and Lau, A. (1998) Quinine for nocturnal leg cramps: a meta-analysis including unpublished data. *Journal of General Internal Medicine*, 13(9), 600-606.

McAlister, F.A., Clark, H.D., van Walraven, C., Straus, S.E., Lawson, F.M., Moher, D. and Mulrow, C.D. (1999) The medical review article revisited: has the science improved? *Annals of Internal Medicine*, 131(12), 947-951.

McAuley, L., Pham, B., Tugwell. P. and Moher, D. (2000) Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet*, 356(9237), 1228-1231.

McDonald, S., Westby, M., Clarke, M., Lefebvre, C. and the Cochrane Centres' Working Group on 50 Years of Randomized Trials. (2002) Number and size of randomized trials reported in general health care journals from 1948 to 1997. *International Journal of Epidemiology*, 31(1), 125-127.

McIntosh, N. (1996) Abstract information and structure at scientific meetings. *Lancet*, 347(9000), 544-545.

Middleton, P. (2003) How allocation concealment is handled in Cochrane reviews. *In: 11th International Cochrane Colloquium, 2003 Oct 26-31, Barcelona, Spain. p. 9.*

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D. and Stroup, D.F. (1999a) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet*, 354(9193), 1896-1900.

Moher, D., Cook, D.J., Jadad, A.R., Tugwell, P., Moher, M., Pham, B. and Klassen, T.P. (1999b) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technology Assessment*, 3(12), 1-98.

Moher, D., Fortin, P., Jadad, A.R., Jüni, P., Klassen, T., Le Lorier, J., Liberati, A., Linde, K. and Penna, A. (1996) Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet*, 347(8998), 363-366.

Moher, D., Jones, A., Lepage, L. and the CONSORT Group (Consolidated Standards of Reporting of Trials) (2001a) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*, 285(15), 1992-1995.

Moher, D., Pham, B., Jones, A., Cook, D.J., Jadad, A.R., Moher, M., Tugwell, P. and Klassen, T.P. (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, 352(9128), 609-613.

Moher, D., Pham, B., Lawson, M.L. and Klassen, T.P. (2003) The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment*, 7(41), 1-90.

Moher, D., Schulz, K.F. and Altman, D.G. (2001b) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet*, 357(9263), 1191-1194.

Montgomery, A.A., Graham, A., Evans, P.H. and Fahey, T. (2002) Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Services Research* [online] 2(1). Available from: www.biomedcentral.com/1472-6963/2/8 [Accessed 22 June 2004].

Mulrow, C.D. (1987) The medical review article: state of the science. *Annals of Internal Medicine*, 106(3), 485-488.

Mulrow, C.D. (1994) Rationale for systematic reviews. *BMJ*, 309(6954), 597-599.

Mulrow, C.D., Thacker, S.B. and Pugh, J.A. (1988) A proposal for more informative abstracts of review articles. *Annals of Internal Medicine*, 108(4), 613-615.

Narine, L., Yee, D.S., Einarson. T.R. and Ilersich, A.L. (1991) Quality of abstracts of original research articles in *CMAJ* in 1989. *Canadian Medical Association Journal*, 144(4), 449-453.

Panush, R.S., Delafuente, J.C., Connelly, C.S., Edwards, N.L., Greer, J.M., Longley, S. and Bennett, F. (1989) Profile of a meeting: how abstracts are written and reviewed. *Journal of Rheumatology*, 16(2), 145-147.

Pearson, K. (1904) Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243-1246.

Petrosino, A., Boruch, R.F., Rounding, C., McDonald, S. and Chalmers, I. (2000) The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions. *Evaluation and Research in Education*, 14(3-4), 206-218.

Pham, B., Platt, R., McAuley, L., Klassen, T.P. and Moher, D. (2001) Is there a "best" way to detect and minimise publication bias? An empirical evaluation. *Evaluation and the Health Professions*, 24(2), 109-125.

Pildal, J. (2004) *Allocation concealment as reported in protocols, published reports and systematic reviews of randomized trials*. Unpublished PhD thesis, University of Copenhagen.

Pitkin, R.M. and Branagan, M.A. (1998) Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA*, 280(3), 267-269.

Pitkin, R.M., Branagan, M.A. and Burmeister, L.F. (1999) Accuracy of data in abstracts of published research articles. *JAMA*, 281(12), 1110-1111.

Ravnskov, U. (1992) Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ*, 305(6844), 15-19.

Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641.

Roy, D., Sankar, V., Hughes, J.P., Jones, A. and Fenton, J.E. (2001) Publication rates of scientific papers presented at the Otorhinolaryngological Research Society meetings. *Clinical Otolaryngology*, 26(3), 253-256.

Rubin, H.R., Redelmeier, D.A., Wu, A.W. and Steinberg, E.P. (1993) How reliable is peer review of scientific abstracts? Looking back at the 1991 Annual Meeting of the Society of General Internal Medicine. *Journal of General Internal Medicine*, 8(5), 255-258.

Sacks, J.J. and Peterson, D.E. (1994) Improving conference abstract selection. *Epidemiology*, 5(6), 636-637.

Scherer, R.W. and Crawley, B. (1998) Reporting of randomized clinical trial descriptors and use of structured abstracts. *JAMA*, 280(3), 269-272.

Scherer, R.W. and Langenberg, P. (2004) Full publication of results initially presented in abstracts (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK, John Wiley & Sons, Ltd.

Schmidmaier, D. (1986) Ask no questions and you'll be told no lies: or how we can remove people's fear of "Grey Literature". *Librarian,* 36(2), 98-112.

Schulz, K.F., Chalmers, I., Hayes, R.J. and Altman, D.G. (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 273(5), 408-412.

Schulz, K.F. and Grimes, D.A. (2002a) Allocation concealment in randomised trials: defending against deciphering. *Lancet*, 359(9306), 614-619.

Schulz, K.F. and Grimes, D.A. (2002b) Blinding in randomised trials: hiding who got what. *Lancet*, 359(9307), 696-700.

Schulz, K.F. and Grimes, D.A. (2002c) Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*, 359(9305), 515-519.

Silagy, C.A., Middleton, P. and Hopewell, S. (2002) Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA*, 287(21), 2831-2834.

Simes, R.J. (1986) Publication bias: the case for an international registry of clinical trials. *Journal of Clinical Oncology*, 4(10), 1529-1541.

Smart, R.G. (1964) The importance of negative results in psychological research. *Canadian Psychologist*, 5(4), 225-232.

Smith, J., Nixon, R., Bueschen, A.J., Venable, D.D. and Henry, H.H. (2002) Impact of blinded versus unblinded abstract review on scientific program content. *Journal of Urology*, 168(5), 2123-2125.

Soares, H.P., Daniels, S., Kumar, A., Clarke, M., Scott, C., Swann, S., Djulbegovic, B. and Radiation Therapy Oncology Group. (2004) Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*, 328(7430), 22-24.

Song, F., Eastwood, A.J., Gilbody, S., Duley, L. and Sutton, A.J. (2000) Publication and related biases. *Health Technology Assessment*, 4(10), 1-115.

Sprague, S., Bhandari, M., Devereaux, P.J., Swiontkowski, M.F., Tornetta, P., Cook, D.J., Dirschl, D., Schemitsch, E.H. and Guyatt, G.H. (2003) Barriers to full-text publication following presentation of abstracts at annual orthopaedic meetings. *Journal of Bone and Joint Surgery American*, 85-A(1), 158-163.

Sterling, T.D., Rosenbaum, W.L. and Weinkam, J.J. (1959) Publication decisions and their effects on inferences drawn from tests of significance - or vice versa. *American Statistical Association Journal*, 54, 30-34.

Stern, J, M. and Simes, R.J. (1997) Publication bias: evidence of delayed publication in a cohort of clinical research projects. *BMJ*, 315(7109), 640-645.

Sterne, J.A., Gavaghan, D. and Egger, M. (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53(11), 1119-1129.

Stewart, L. and Tierney, J. (2000) Publication bias is a serious threat to the validity of Cochrane reviews. *In: Eighth International Annual Cochrane Colloquium, 2000 Oct 25-29, Cape Town, South Africa.* p.8.

STN International. *SIGLE (System for Information on Grey Literature in Europe): database summary sheet* [online] STN International, Chemical Abstracts Service, Columbus, Ohio, USA. Available from: www.cas.org/ONLINE/DBSS/Sigless.Html [Accessed 5 December 2003].

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A. and Thacker, S.B. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. *JAMA*, 283(15), 2008-2012.

Sutton, A.J., Abrams K.R., Jones, D.R. and Sheldon, T.R. (2000) *Methods for Meta-Analysis in Medical Research.* Chichester, John Wiley & Sons Ltd.

Sykes, R. (1998) Being a modern pharmaceutical company: involves making information available on clinical trial programmes. *BMJ*, 317(7167),1172.

Taddio, A., Pain, T., Fassos, F.F., Boon, H., Ilersich, A.L. and Einarson T.R. (1994) Quality of non-structured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association. Canadian Medical Association Journal*, 150(10), 1611-1615.

Taus, C., Pucci, E., Giuliani, G., Telaro, E. and Pistotti, V. (1999) The use of "grey literature" in a sub-set of neurological Cochrane reviews. *In: Seventh International Annual Cochrane Colloquium, 1999 Oct 5-9, Rome, Italy.* p.30.

Timmer, A., Sutherland, L.R. and Hilsden, R.J. (2003) Development and evaluation of a quality score for abstracts. *BMC Medical Research Methodology* [online] 3(2). Available from: www.biomedcentral.com/1471-2288/3/2 [Accessed 22 June 2004].

Tonks, A. (2002) A clinical trials register for Europe. *BMJ*, 325(7376), 1314-1315.

Tooher, R., Middleton, P., Griffin, T., Pham, C. and Hopewell, S. (2004) How different are conference abstracts of surgical RCTs from their subsequent full publication? *In:*

12th International Cochrane Colloquium, 2004 Oct 2-6, Ottawa, Canada (forthcoming).

Tramer, M.R., Reynolds, D.J., Moore, R.A. and McQuay, H.J. (1997) Impact of covert duplicate publication on meta-analysis: a case study. *BMJ*, 315(7109), 635-640.

Usatoff, V. and Waxman, B.P. (1997) A critical evaluation of free paper abstracts accepted for the 1996 RACS Annual Scientific Congress. *Australian and New Zealand Journal of Surgery*, 67(1), 45-46.

Van Loo, J. (1985) Medical and psychological effects of unemployment: a 'grey' literature search. *Health Libraries Review*, 2(2), 55-62.

Vickers, A.J. and Smith, C. (2000) Incorporating data from dissertations in systematic reviews. *International Journal of Technology Assessment in Health Care*, 16(2), 711-713.

Vilstrup, H. and Sorensen, H.T. (1998) A comparative study of scientific evaluation of abstracts submitted to the 1995 European Association for the Study of the Liver Copenhagen meeting. *Danish Medical Bulletin*, 45(3), 317-319.

Von Elm, E., Costanza, M.C., Walder, B. and Tramer, M.R. (2003) More insight into the fate of biomedical meeting abstracts: a systematic review. *BMC Medical Research Methodology* [online] 3(12). Available from: www.biomedcentral.com/1471-2288/3/12 [Accessed 22 June 2004].

Weber, E.J., Callaham, M.L., Wears, R.L., Barton, C. and Young, G. (1998) Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA*, 280(3), 257-259.

Weintraub, W.H. (1987) Are published manuscripts representative of the surgical meeting abstracts? An objective appraisal. Journal of *Pediatric Surgery*, 22(1), 11-13.

White, K.R. (1982) The relation between socio-economic status and academic achievement. *Psychological Bulletin*, 91(3), 461-481.

Wormald, R., Bloom, J., Evans, J. and Oldfield, K. (1997) Publication bias in eye trials. *Second International Conference Scientific Basis of Health Services and Fifth International Cochrane Colloquium, 1997, Oct 5-8, Amsterdam, The Netherlands.* p.230.

APPENDIX 1: Publications relating to this thesis

This appendix contains a bibliographic list of publications resulting from the work of this thesis.

A selection of these publications details below, which have been published as journal articles are also contained within this appendix:

- Hopewell, S. (2003) Assessing the impact of abstracts from the Thoracic Society of Australia and New Zealand in Cochrane reviews. *Respirology*, 8(4), 509-512.

- Hopewell, S. and McDonald, S. (2003) Full publication of trials initially reported as abstracts in the *Australian and New Zealand Journal of Medicine* 1980-2000. *Internal Medicine Journal*, 33(4), 192-194.

- Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2004) Grey literature in meta-analyses of randomized trials of health care interventions (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

Publications resulting from the work of this thesis

2002

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2002) Grey literature in meta-analyses of randomized trials - a systematic review. *In: Fourth Symposium on Systematic Reviews: Pushing the Boundaries, 2002 Jul 2-4, Oxford, UK.* p. 44.

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2002) Grey literature in meta-analyses of randomized trials - a systematic review. *In: 10th International Cochrane Colloquium, 2002 Jul 31 - Aug 3, Stavanger, Norway.* p. 29.

Mallett, S., Hopewell, S. and Clarke, M. (2002) The use of grey literature in the first 1000 Cochrane reviews. *In: Fourth Symposium on Systematic Reviews: Pushing the Boundaries, 2002 Jul 2-4, Oxford.* p.43.

2003

Hopewell, S. (2003) Assessing the impact of abstracts from the Thoracic Society of Australia and New Zealand in Cochrane reviews. *Respirology*, 8(4), 509-512.

Hopewell, S. and Clarke, M. (2003) Quality of trials reported as conference abstracts: how well are they reported? *In: 11th International Cochrane Colloquium, 2003 Oct 26-31, Barcelona, Spain.* p. 38.

Hopewell, S., MacKinnon, L. and Clarke, M. (2003) Conference abstracts that contain "possible" randomized trials - are they worth finding? *In: 11th International Cochrane Colloquium, 2003 Oct 26-31, Barcelona, Spain.* p. 34.

Hopewell, S. and McDonald, S. (2003) Full publication of trials initially reported as abstracts in the Australian and New Zealand Journal of Medicine 1980-2000. *Internal Medicine Journal*, 33(4), 192-194.

2004

Hopewell, S. (2004) Trials reported as abstracts and full publications: how do they compare? *Clinical Trials*, 1 (Suppl 1), 216-217.

Hopewell, S., Clarke, M. and Askie, L. (2004) Trials reported as abstracts and full publications: how do they compare? *In: 12th International Cochrane Colloquium, 2003 Oct 2-6, Ottawa, Canada* (forthcoming).

Hopewell, S., Clarke, M. and Mallett, S. (2004) Grey literature in systematic reviews. In: Publication bias in meta-analysis: Prevention, Assessment and Adjustment. M. Rothstein, A. Sutton, M. Borenstein (editors). Chichester, Son Wiley & Sons Ltd. (forthcoming)

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2004) Grey literature in meta-analyses of randomized trials of health care interventions (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

Tooher, R., Middleton, P., Griffin, T., Pham, C. and Hopewell, S. How different are conference abstracts of surgical RCTs from their subsequent full publication? *In: 12th International Cochrane Colloquium, 2003 Oct 2-6, Ottawa, Canada* (forthcoming).

Internal Medicine Journal 2003; 33: 192–194

BRIEF COMMUNICATION

Full publication of trials initially reported as abstracts in the *Australian and New Zealand Journal of Medicine* 1980–2000

S. HOPEWELL[1] and S. McDONALD[2]

[1]*UK Cochrane Centre, Oxford, United Kingdom and* [2]*The Australasian Cochrane Centre, Monash Institute of Health Services Research, Melbourne, Australia*

Abstract

Failure to identify all relevant reports of controlled trials is a potential source of bias in systematic reviews of health-care interventions. The present study aims to identify how many reports of trials initially published as conference abstracts in the *Australian and New Zealand Journal of Medicine* were subsequently published in full. We identified trial reports by handsearching conference abstracts published in the Journal from 1980 to 2000. We then searched the *Cochrane Controlled Trials Register* and PubMed to determine how many of these had been subsequently published in full. A total of 962 reports of controlled trials was identified from the conference proceedings of 17 medical societies. Of these, 589 (61%) reports of trials were subsequently published in full, and on average within 1–2 years. Handsearching conference abstracts identified a large number of reports of controlled trials, over one-third of which were unpublished and therefore not easily accessible. (Intern Med J 2003; 33: 192–194)

Key words: handsearching, publication bias, trial identification.

INTRODUCTION

The Cochrane Collaboration is an international organization that aims to help people make well-informed decisions about health care by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of health-care interventions.[1] The validity of the results of a systematic review is highly dependent on the results of the underlying data and, as such, requires the identification of as unbiased and complete a set of relevant studies as possible.

The results of scientific research are often presented at professional meetings and summarized in abstracts. These abstracts are often available only in proceedings or journal supplements and are rarely indexed in databases such as MEDLINE. Research suggests that just over half of all conference abstracts describing the results of clinical trials are published in full, and those with significant results are published in full more frequently than those with non-significant results.[2] If a systematic review only includes data from those reports of research that are published in full, then the results and conclusions of that review may be biased.

The Cochrane Collaboration, through its considerable efforts to identify and register reports of controlled trials, is increasing the pool of eligible studies available to reviewers and thereby helping to minimize the biases just described. In addition to systematic searches of databases like MEDLINE and EMBASE, many hundreds of journals and conference proceedings have been systematically searched by hand for reports of trials.[3] *The Cochrane Central Register of Controlled Trials* (formerly known as *The Cochrane Controlled Trials Register*) acts as the central collection point for all these efforts, containing over 330 000 reports, and serves as the best available resource for those searching for randomized trials of health-care interventions.[4]

Correspondence to: Sally Hopewell, The UK Cochrane Centre, Summertown Pavilion, Middle Way, Oxford, OX2 7LG, United Kingdom. Email: shopewell@cochrane.co.uk

Received 16 May 2002; accepted 2 September 2002.

As part of this international effort, the *Australian and New Zealand Journal of Medicine* was handsearched to identify reports of controlled trials. The Journal, which was first published in 1952 as the *Australasian Annals of Medicine*, changed to the *Australian and New Zealand Journal of Medicine* in 1971, and began including the proceedings of several medical societies in 1980. In 2001 the Journal was renamed the *Internal Medicine Journal*.

The aim of this study was to handsearch all conference abstracts published in the *Australian and New Zealand Journal of Medicine* to identify reports of controlled trials, and to determine how many of these had subsequently been published in full.

METHODS

Handsearching

The abstracts of all conference proceedings published in the *Australian and New Zealand Journal of Medicine* (1980–2000) were handsearched to identify all reports of randomized controlled trials (RCT), as defined by the Cochrane Collaboration. This definition requires that 'the participants (or other units) were definitely or possibly assigned to one of two (or more) alternative forms of health care using a process of random allocation'. Reports of trials that used a quasi method of randomization (such as alternation, date of birth or case record number), or where randomization could not be ruled out, were coded as controlled clinical trials (CCT).

All reports judged to be potentially relevant by the handsearcher were then individually assessed by one of the authors to confirm that they met the Cochrane eligibility criteria.

Assessing full publication

We assessed each RCT and CCT identified from among the conference abstracts to determine whether it had been published as a full report. Issue 1, 2002 of *The Cochrane Controlled Trials Register* was used to search for the first author listed in the abstract (and then second or third author if this was not successful). Where a large number of 'hits' was retrieved, a key word from the title was used to help refine the search. We then examined the publications retrieved to ensure that the full report corresponded with the abstract. We then repeated the searches in PubMed (1966 to May 2002) for those abstracts not identified as having been published as a full report in *The Cochrane Controlled Trials Register*. Any uncertainties were checked and resolved by discussion. Because *The Cochrane Controlled Trials Register* contains nearly 150 000 reports of trials from sources other than MEDLINE (including 40 000 reports unique to EMBASE), we decided not to extend the search to other databases. The time interval between the year of the conference at which the trial reported in the abstract had been presented, and full publication of its results, was also recorded.

RESULTS

We identified a total of 962 reports of controlled trials by searching the conference abstracts published in the *Australian and New Zealand Journal of Medicine* between 1980 and 2000. Of these, 726 (76%) were reports of RCT and 236 (24%) were reports of CCT. The median number of trials published per year was 43 (inter-quartile range 27.5–58.5). We identified trials from the conference proceedings of 17 medical societies with the highest yield coming from the Cardiac Society of Australia and New Zealand (273 RCT; 65 CCT) (Table 1).

Sixty-one per cent ($n = 589$) of the trials initially presented as abstracts were subsequently published as full reports (Table 1). This varied between medical societies, from 46% for the Australian and New Zealand Society of Nuclear Medicine to 79% for the Australian Society of Neurologists. Thirty-one of the 589 trials that we were able to identify full publications for were retrieved by searching PubMed and were not identifiable in the *Cochrane Controlled Trials Register*. As of May 2002, the remaining 373 (39%) abstracts remained unpublished. The average time from presentation at the conference to full publication of results was 1–2 years, ranging from less than 1 year to 10 years. Eighty-four per cent were published within 2 years; 95% were published within 4 years; and 99% were published within 8 years. Time to publication was measured from the year the abstract was presented at the conference to the year of full publication of its results. Four trials had already been published in full 1 year before presentation as an abstract at a conference.

A large number of controlled trials, from a number of different medical specialities, was identified by handsearching conference abstracts published in the *Australian and New Zealand Journal of Medicine*. Only 61% of the trials reported in these abstracts were subsequently published as full reports and 84% of those were published within 2 years. The remaining abstracts were unpublished and therefore could not easily have been identified by other means. This

Table 1 Number of trials reported in abstracts and full publication of their results

Medical society*	Number trials	Published in full (%)
Cardiac Society of Australia and New Zealand	338	221 (65)
Thoracic Society of Australia and New Zealand	187	101 (54)
Gastroenterological Society of Australia	138	89 (64)
Australian Rheumatology Association	100	57 (57)
Haematology Society of Australia	42	23 (55)
Australian Society for Geriatric Medicine	38	27 (71)
Australian and New Zealand Society of Nuclear Medicine	28	13 (46)
Australian Association of Neurologists	19	15 (79)
Australasian Society for Infectious Diseases	18	13 (72)
Australian Society of Nephrology	17	9 (53)
Australian Diabetes Society	14	8 (57)
The Royal Australasian College of Physicians	11	8 (73)
Others	12	6 (50)
Total	962	589 (61)

*The name of the medical society may have changed over time.

figure is consistent with other studies showing that just over half of all conference abstracts describing the results of clinical trials are subsequently published in full.[2] Had we broadened our search to include other bibliographic databases, it is possible that we may have identified more full reports, although we believe the number would have been small. It is also possible that a small number of trials presented as conference abstracts in the year 2000 was subject to a time-lag to full publication. Again, we do not expect this to have substantially affected the results of this study.

Our study supports the Cochrane Collaboration's international effort of handsearching often-inaccessible material for reports of controlled trials in order to enable authors of systematic reviews to identify as complete a set of relevant studies as possible. A recent study showed that 20% of Cochrane reviews contain trials for which the only published source is a conference abstract.[5] Having identified large numbers of trials published only as conference abstracts, further research is required to determine the impact that these trials have on the overall results of systematic reviews.[6]

ACKNOWLEDGEMENTS

The present study was carried out at the Australasian Cochrane Centre, which is funded by the Department of Health and Ageing in Australia and the UK Cochrane Centre, which is funded by the National Health Service Research and Development Programme.

The authors would like to thank Marian Bare and Marie Misso for reading and coding the abstracts, and Sarah Hetrick for help with carrying out the electronic searches. Thanks are also due to Mike Clarke and Sally Green for their helpful comments on drafts of this manuscript.

The views expressed in this paper represent those of the authors and are not necessarily the views or the official policy of The Cochrane Collaboration.

REFERENCES

1 Bero L, Rennie D, The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of healthcare. JAMA 1995; 274: 1962–4.
2 Scherer RW, Langenberg PFull publication of results initially presented in abstracts. (Cochrane Methodology Review). In: The Cochrane Library, Issue 4, 2002. Oxford: Update Software.
3 Lefebvre C, Clarke M. Identifying randomised trials. In: Egger, M, Davey Smith, G, Altman, D, eds. Systematic Reviews in Health Care: Meta-Analysis in Context. London: BMJ Publishing; 2001; 69–86.
4 Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S and the CENTRAL Development Group. Development of the Cochrane Collaboration's central register of controlled trials. Evaluation Health Professions 2002; 25: 38–64.
5 Mallett S, Hopewell S, Clarke M. The use of Grey Literature in the First 1000 Cochrane Reviews. In: Proceedings of the 4th Symposium on Systematic Reviews: Pushing the Boundaries; 2002 July 2–4; 2002; Oxford; Symposium on Systematic Reviews; 2002; .
6 Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials (Protocol for a Cochrane Review). In: The Cochrane Library, Issue 4, 2002. Oxford: Update Software.

Respirology (2003) **8**, 509–512

ORIGINAL ARTICLE

Assessing the impact of abstracts from the Thoracic Society of Australia and New Zealand in Cochrane reviews

SALLY HOPEWELL

UK Cochrane Centre, Oxford, United Kingdom

Assessing the impact of abstracts from the Thoracic Society of Australia and New Zealand in Cochrane reviews
HOPEWELL S. *Respirology* 2003; **8**: 509–512
Objective: The aim of this study was to assess the potential impact of including trials, reported in conference abstracts from the Thoracic Society of Australia and New Zealand, in Cochrane reviews.
Methodology: Abstracts from the Thoracic Society of Australia and New Zealand, published in the *Australian and New Zealand Journal of Medicine* (1981–1998), were read to identify all reports of randomized trials. A search was carried out of the *Cochrane Database of Systematic Reviews* (Issue 1, 2002) for each trial reported in a conference abstract to try to identify Cochrane reviews in which the conference abstract might be eligible for inclusion. If it was unclear, the authors of the review were contacted.
Results: A total of 187 reports of randomized trials were identified: 101 (54%) had been published as a full report and 86 (46%) remained unpublished. Thirty-four (72%) were reports of randomized controlled trials and 52 (28%) were quasi-randomized or possibly randomized trials. The total number of patients included in the trials was 9691; range 4–1203 (median 20; IQR 11–47). No possible Cochrane review was found for 145 of the 187 trials reported in the conference abstracts. Possible reviews were identified for 42 trials, 24 of which were already mentioned in Cochrane reviews. For the remaining 18 trials, only three were said to be eligible for inclusion.
Conclusion: A search of conference abstracts identified a number of reports of randomized trials, potentially eligible for inclusion in reviews of health care. However, the majority of trials were not relevant for inclusion in an existing Cochrane review. This is most likely because there are currently too few reviews to deal with the topics covered in the abstracts.

Key words: publication bias, trial identification, trial quality.

INTRODUCTION

The Cochrane Collaboration is an international organization that aims to help people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions. The validity of the results of a systematic review is highly dependent on the results of the underlying data and, as such, requires the identification of as unbiased and complete a set of relevant studies as possible.

There is now much empirical evidence to demonstrate that studies with significant results are more likely to be published than those with non-significant results,[1] and that studies with significant results are often published earlier than those with non-significant results.[2] There is also evidence to suggest that only about half of all conference abstracts describing the results of clinical trials are published in full, and that those with significant results are published in full more frequently than those with non-significant results.[3]

Therefore, if only published reports of research are included in a systematic review, the results and conclusions of that review may be biased. The inclusion of grey literature (for example conference abstracts, research reports, dissertations, policy documents and personal correspondence) in systematic reviews may help to overcome some of the problems of publica-

Correspondence: Sally Hopewell, The UK Cochrane Centre, Summertown Pavilion, Middle Way, Oxford OX2 7LG, UK. Email:shopewell@cochrane.co.uk

The views expressed in this paper represent those of the author and are not necessarily the views or the official policy of The Cochrane Collaboration.

Received 9 October 2002; revised 26 March 2003; accepted for publication 31 March 2003.

tion bias, which can arise due to the selective availability of data. There is also now some evidence to suggest that the exclusion of grey literature from a systematic review can lead to an exaggeration of the effect of treatment.[4]

However, the identification of relevant studies in the grey literature can be particularly time-consuming and difficult. Even when identified, the inclusion of grey literature in a systematic review can still be difficult due to the limited and often incomplete amount of information included.[5]

OBJECTIVE

The aim of this study was to assess the potential impact of including a sample of grey literature in Cochrane reviews published in *The Cochrane Database of Systematic Reviews* in *The Cochrane Library* (Issue 1, 2002).

The sample of grey literature comprised abstracts presented at the annual conference of the Thoracic Society of Australia and New Zealand between 1981 and 1998 and published in the *Australian and New Zealand Journal of Medicine* (continued as the *Internal Medicine Journal* from 2001). This sample was selected because of the high number of trials reported in the conference abstracts and the expected high number of potentially eligible Cochrane reviews.

METHODS

Handsearching

As part of a larger project, all conference abstracts published in the *Australian and New Zealand Journal of Medicine* (1980–1998)[6] were read to identify all reports of randomized trials, as defined by the Cochrane Collaboration. This definition requires that the participants (or other units) were definitely or possibly assigned prospectively to one of two (or more) alternative forms of health care using a process of random allocation. For the purposes of the present study only the 187 reports of randomized trials presented as abstracts at the annual conference of the Thoracic Society of Australia and New Zealand were assessed. These abstracts primarily included reports of randomized trials in asthma, chronic lung disease, croup, bronchitis and cystic fibrosis.

Assessing full publication

Each conference abstract was assessed to determine if it had been published as a full report. *The Cochrane Controlled Trials Register (CENTRAL)* in *The Cochrane Library* (Issue 1, 2002) was used to search for the first author (and then second author if this was not successful) listed in the abstract. Where a large number of hits were retrieved a key word from the conference abstract was used to help refine the search. The search was then repeated using the US National Library of Medicine's PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) for those conference abstracts for which a full report was not found in *CENTRAL* (March 2002). Possible publications were then examined to see if the full report corresponded with the subject matter in the conference abstract. The time interval between the year of the conference at which the relevant conference abstract had been presented, and full publication of the trial's results was also recorded.

Assessment of methodological quality

The methodological quality of each trial reported in a conference abstract was assessed based on the method of allocation concealment, which is the process used to prevent foreknowledge of group assignment in the randomized controlled trial. Each trial reported in a conference abstract was assessed as having either adequate allocation concealment (e.g. central randomization office) or inadequate (e.g. date of birth or case record number) or unclear allocation concealment. This method of assessing trial quality was used because it is the method most commonly used by Cochrane reviewers.

Searching for possible Cochrane reviews

A search was carried out of the *Cochrane Database of Systematic Reviews* (Issue 1, 2002), for each trial reported in a conference abstract, to try to identify Cochrane reviews in which the conference abstract might be eligible for inclusion. This was done by reading each conference abstract and then using the search function in *The Cochrane Library* to search for key words from the conference abstract. Retrieved reviews were then assessed by reading the title and abstract of the review, to determine whether the trial might be relevant for inclusion.

Where a possible review was identified, its list of references was checked to determine if the trial reported in the conference abstract was already included in the review, either as an abstract or as a full report. If the trial was not already included in the review, I then assessed if it met the inclusion criteria for the review. This was done using the 'criteria for considering studies' section of the review, and by assessing the following eligibility criteria: (i) types of studies (e.g. the study design); (ii) types of participants (e.g. the disease or condition of interest and the population and setting); (iii) types of interventions (e.g. the intervention of interest and what it was being compared against); and (iv) types of outcome measures (e.g. the specific outcomes which were being assessed).

If it was unclear whether the trial reported in the conference abstract met the full eligibility criteria for inclusion in a review, I wrote to the corresponding author of the review, enclosing a copy of the abstract, and asked for his/her help in determining its eligibility.

RESULTS

Full publication

A total of 187 reports of randomized trials were identified in the abstracts from the annual conferences of the Thoracic Society of Australia and New Zealand (1981–1998). Of these, 101 (54%) had been published as a full report. The average time to full publication of the trials' results was 1–2 years, ranging from less than 1 year up to 6 years. As of May 2002, 86 (46%) reports of randomized trials presented as conference abstracts remained unpublished.

Design

Of the 187 reports of randomized trials identified, 134 (72%) were reports of randomized controlled trials and 52 (28%) were reports of quasi-randomized or possibly randomized trials.

Number of patients

A total of 181 abstracts included information on the number of randomized patients. The total number of patients included in the trials was 9691, ranging between four and 1203 patients per trial. The median number of patients per trial was 20 (IQR 11–47). This was similar for those trials reported in conference abstracts that had subsequently been published as a full report (median 20; IQR 12–48) and was slightly more than that for trials published only as conference abstracts (median 16; IQR 10–46) (Table 1).

Assessment of methodological quality

The method of allocation concealment could be determined in only four of the trials reported in the conference abstracts. It was not possible to determine the method of allocation concealment used in the majority of the trials because not enough information was provided in the abstract.

Identifying possible Cochrane reviews

No possible Cochrane review was found for 145 of the 187 trials reported in the conference abstracts (Table 2). Possible reviews were identified for 42 trials, 24 of which were already mentioned in Cochrane reviews. Twenty-two trials were referenced as reported in the full publication (four of which were listed as an 'excluded study' by the reviewers) and two trials were referenced as reported in the conference abstract.

Table 1 Characteristics of the trials reported in the abstracts

	Overall	Published in full	Published as abstracts
Total no. abstracts	187	101	86
Year of publication of abstracts	1981–1998	1981–1998	1981–1998
Number of randomized controlled trials	134	76	58
Number of CCTs[a]	53	25	28
Total no. patients[b]	9691	5903	3788
Median no. patients (IQR)	20 (IQR 11–47)	20 (IQR 12–48)	16 (IQR 10–46)
Range of patients	4–1203	5–1203	4–370
Allocation concealment	Adequate = 4 Unclear = 183	Adequate = 3 Unclear = 98	Adequate = 1 Unclear = 85

[a]CCTs refers to those reports of trials that used a quasi method of randomization (such as alternation, date of birth or case record number), or where randomization could not be ruled out, which were coded as controlled clinical trials.

[b]Data for the number of patients is given for 181 abstracts; six abstracts did not include this information (two published and four unpublished).

Table 2 Trials reported in conference abstracts for which there was a possible Cochrane review

	Overall (n = 187)	Published in full (n = 101)	Published as abstracts (n = 86)
Possible review identified	42	28	14
(Already in review)	24	22[a]	2
(Said to be eligible by reviewer)	3	0	3
(Said to be ineligible by reviewer)	15	6	9
No review identified	145	73	72

[a]Four abstracts published in full were initially thought to be eligible for inclusion in the review but had been listed as an excluded study by the authors of the review.

For the remaining 18 trials reported in conference abstracts, it was unclear if they met the full eligibility criteria for the review and the author of each review was contacted for help. If this was not possible an expert in the field was contacted for advice. Only three of the 18 trials reported in conference abstracts were said to be eligible for inclusion in a review (one author said they would need to contact the authors of the trial for clarification); the remaining 15 trial reports were said to be ineligible for inclusion. The reasons given as to why the trial reports were not eligible for inclusion in the review included: no clinical data available (one); the types of participants were different to those assessed in the review (one); and the trial compared an intervention which was not assessed in the review (six).

CONCLUSIONS

A search of the abstracts from the annual conferences of the Thoracic Society of Australia and New Zealand identified a number of reports of randomized trials, which were potentially eligible for inclusion in systematic reviews of healthcare interventions. Just over half of these trials reported as abstracts had subsequently been published as a full report and trials reported in these abstracts were slightly larger than those that remained unpublished.

The method of allocation concealment was unclear and poorly reported in all but four of the 187 trials reported in the conference abstracts. The poor reporting of methodological quality in conference abstracts is a problem routinely faced by reviewers who then need to contact the authors of the trial for more information.[7] Furthermore, six of the 187 trials did not include data on the number of participants included in the trial. Improved reporting of conference abstracts, for example including a description of the methods of allocation concealment and blinding, is needed.

Twenty-four of the 187 trials, reported in conference abstracts, were already included in Cochrane reviews, the majority of which referenced the trial as reported in the full publication. Only three trials were identified, which were suitable for inclusion but were currently missing from a Cochrane review, all of which were only published in conference abstracts. Furthermore, one of the authors of the review said they would still need to contact the authors of the trial report for further clarification. The majority of trials identified were not judged relevant for inclusion in an existing Cochrane review. This is most likely to be because there are currently too few Cochrane reviews to deal with the topics covered in the abstracts.

This study is based on a relatively small sample of trials within a specific area of healthcare and it is therefore possible that the results of this study may not be the same across other areas of healthcare. However, a recent study predicts that at least another 10 000 Cochrane reviews are needed to cover a substantial proportion of the trials, which have already been identified in healthcare.[8]

ACKNOWLEDGEMENTS

I am grateful to those authors of Cochrane reviews who provided clarification on whether a trial reported in an abstract was eligible for inclusion in their review. The author would also like to thank Mike Clarke and Anne Lusher for their helpful comments on drafts of this manuscript.

REFERENCES

1 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol. Assess.* 2000; **4**: 1–115.

2 Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials (Cochrane Methodology Review). In: *The Cochrane Library*. Update Software, Oxford, Issue 3, 2003.

3 Scherer RW, Langenberg P. Full publication of results initially presented in abstracts (Cochrane Methodology Review). In: *The Cochrane Library*. Update Software, Oxford, Issue 3, 2003.

4 Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of healthcare interventions (Cochrane Methodology Review). In: *The Cochrane Library*. Update Software, Oxford, Issue 3, 2003.

5 Cook DJ, Guyatt GH, Ryan G *et al.* Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993; **269**: 2749–53.

6 Hopewell S, McDonald S. Full publication of trials initially reported as abstracts in the Australian and New Zealand Journal of Medicine 1980–2000. *Intern. Med. J.* 2003; **33**: 192–4.

7 Clarke M, Greaves L. Searching meeting abstracts: a comparison between an experienced handsearcher and an indexed, electronic database. In: *4th Cochrane Colloquium*, Adelaide, Australia, October, 1996.

8 Mallett S, Clarke M. A typical Cochrane review: how many more are needed to cover existing evidence? In: *10th Cochrane Colloquium*, Stavanger, Norway, July, 2002.

UPDATE SOFTWARE

Grey literature in meta-analyses of randomized trials of health care interventions

Hopewell S, McDonald S, Clarke M, Egger M.

This is a reprint of a Cochrane methodology review, prepared and maintained by the Cochrane Collaboration and published in *The Cochrane Library* 2003, Issue 2



The Cochrane Library is available by subscription from:
Update Software Ltd, Summertown Pavilion, Middle Way, Oxford OX2 7LG, UK
Tel: +44 (0)1865 513902 Fax: +44 (0)1865 516918
E-mail: info@update.co.uk
Website: http://www.update-software.com

Update Software Inc, 1070 South Santa Fe Avenue, Suite 21, Vista, CA 92084, USA
Tel: +1 760 631 5844 Fax: +1 760 631 5848
E-mail: info@cochranelibrary.com
Website: http://www.cochranelibrary.com

© Update Software 1996 - 2003
All rights reserved. This document may not be reproduced or published elsewhere, in whole or in part, without the written consent of Update Software Ltd.

TABLE OF CONTENTS

ABSTRACT....1
BACKGROUND....2
OBJECTIVES....2
CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW....2
SEARCH STRATEGY FOR IDENTIFICATION OF STUDIES....2
METHODS OF THE REVIEW....3
DESCRIPTION OF STUDIES....3
METHODOLOGICAL QUALITY....4
RESULTS....4
DISCUSSION....5
REVIEWER'S CONCLUSIONS....5
ACKNOWLEDGEMENTS....6
POTENTIAL CONFLICT OF INTEREST....6
SOURCES OF SUPPORT....6
REFERENCES....7
TABLES....8
Characteristics of included studies....8
Table 01 Characteristics of the meta-analyses....10
Table 02 Characteristics of the trials included in meta-analyses....11
Table 03 Language of publication of the trials....12
Table 04 Type of grey literature....12
Table 05 Methodological quality of the randomized trials included in the meta-analyses....13
Table 06 Summary outcome measures for grey and published trials....13
COVER SHEET....15

Copyright © Update Software Ltd.

Grey literature in meta-analyses of randomized trials of health care interventions

Hopewell S, McDonald S, Clarke M, Egger M.

This Review should be cited as:
Hopewell S, McDonald S, Clarke M, Egger M.. Grey literature in meta-analyses of randomized trials of health care interventions (Cochrane Methodology Review) In: *The Cochrane Library*, Issue 2, 2003. Oxford: Update Software.
Date of most recent amendment: 18 November 2002
Date of most recent substantive amendment: 12 July 2002

ABSTRACT

Background
The inclusion of grey literature (i.e. literature that has not been formally published) in systematic reviews may help to overcome some of the problems of publication bias, which can arise due to the selective availability of data.

Objectives
To review systematically research studies, which have investigated the impact of grey literature in meta-analyses of randomized trials of health care interventions.

Search strategy
We searched the Cochrane Methodology Register (The Cochrane Library Issue 1, 2002), MEDLINE (1966 to February 2002), the Science Citation Index (April 2002) and contacted researchers who may have carried out relevant studies.

Selection criteria
A study was considered eligible for this review if it compared the effect of the inclusion and exclusion of grey literature on the results of meta-analyses of randomized trials.

Data collection and analysis
Data were extracted from each report independently by two reviewers. The main outcome measure was an estimate of the impact of trials from the grey literature on the pooled effect estimates of the meta-analyses. Information was also collected on the area of health care, the number of meta-analyses, the number of trials, the number of trial participants, the year of publication of the trials, the language and country of publication of the trials, the number and type of grey and published literature, and methodological quality.

Main results
Eight studies meeting the inclusion criteria were identified. Four studies contained multiple meta-analyses and four contained single meta-analyses. Of the included studies, four multiple and three single meta-analyses, found that published trials showed an overall greater treatment effect than grey trials. This difference was statistically significant in two of the four multiple meta-analyses. The remaining single meta-analysis found that published trials showed no effect of treatment and that grey trials showed a negative treatment effect; this difference was not statistically significant. Overall there were more published trials included in the meta-analyses than grey trials (median 46 (IQR 4-300) versus 5.5 (IQR 4-88)). Published trials had more participants on average. In the two studies that assessed methodological quality of the included trials, the published trials were of higher quality than the grey trials. The most common types of grey literature were abstracts (49%) and unpublished data (33%).

Reviewers' conclusions
This review suggests that published trials are generally larger and may show an overall greater treatment effect than grey trials. This has important implications for reviewers who need to ensure they identify grey trials, in order to minimise the risk of introducing bias into their review.

Copyright © Update Software Ltd.

BACKGROUND

The validity of a systematic review is highly dependent on the results of the underlying data. The inclusion of grey literature in systematic reviews may help to overcome some of the problems of publication bias, which can arise due to the selective availability of data. The definition of what constitutes grey literature varies (Loo 1985; Alberani 1990; Cook 1993; McAuley 2000; Song 2000; British Library; SIGLE). However, it is generally assumed to include literature that has not been formally published, has limited distribution or is not available through conventional channels (Auger 1998). Examples of grey literature include conference abstracts, research reports, book chapters, unpublished data, dissertations, policy documents and personal correspondence.

In social sciences research, it has been known for some time that omitting unpublished studies from a meta-analysis can magnify the effect of the intervention. For example, in a study of nine meta-analyses published between 1976 and 1980, the average experimental effects found in studies published in journals was larger than the corresponding effect from theses and dissertations (Glass 1981). There is now some evidence in support of this in health research, suggesting that the exclusion of grey literature from meta-analyses can lead to an exaggeration of the effect of treatment (McAuley 2000; Sterne 2000).

However, the identification of relevant studies in the grey literature and their inclusion in systematic reviews can be particularly time-consuming and difficult. There is also some controversy as to whether unpublished studies should be included in meta-analyses because they might be incomplete and their methodological quality can be difficult to assess. A survey by Cook and colleagues in the early 1990s showed that 78% of meta-analysts and methodologists felt that unpublished material should be included in meta-analyses compared to only 47% of journal editors (Cook 1993).

Research is needed to help assess the potential implications for reviewers of not including grey literature in systematic reviews of health care interventions. This review aims to build on work carried out by one of the authors (Egger 2001) and systematically review research studies that have investigated the impact of including grey literature in meta-analyses of health care interventions.

OBJECTIVES

To review systematically empirical studies, which have investigated the impact of including grey literature in meta-analyses of randomized trials of health care interventions.

CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW

Types of studies

A research study was considered eligible for inclusion in this review if it compared the impact of the inclusion and exclusion of grey literature on the results of one or more meta-analyses of randomized trials. The definition of what constitutes grey literature was that used by the authors in each of the empirical studies. A meta-analysis was defined as being where the results of two or more trials are calculated and then statistically pooled to produce a single estimate of treatment effect.

Types of data

Information was collected on the area of health care, the number of meta-analyses, the number of randomized trials, the number of participants in these trials, the year of publication of the trials, the language of publication of the trials, the number and type of grey and published literature, and the methodological quality of the empirical studies and the randomized trials they included.

Types of comparisons

The inclusion and the exclusion of trials from the grey literature on the pooled effect estimate of the meta-analysis.

Types of outcome measures

The main outcome measure was an estimate of the impact of trials from the grey literature on the pooled effect estimates of the meta-analyses. We had hoped to do this by calculating the ratio of the odds ratio between the results of the grey literature and the results of the published literature and estimating the percentage change. However, the summary outcome measures were frequently reported differently in each of the empirical studies. We therefore had to use the summary outcome measure reported by the authors of the included studies.

SEARCH STRATEGY FOR IDENTIFICATION OF STUDIES

Studies were sought from the Cochrane Methodology Register (The Cochrane Library Issue 1, 2002) and MEDLINE (1966 to February 2002). The following search strategy was run against MEDLINE on OVID.

1 (Meta-Analysis or Clinical-Trials or Randomized-Controlled-Trials or Selection-Bias or Review-Literature).sh
2 (meta-analys$ or systematic review$ or data synthes$).tw
3 or/1-2
4 (grey literature or gray literature or unpublished).tw
5 exp Publication Bias
6 or/4-5
7 3 and 6

Studies were also sought during the handsearching of selected journals, which is being carried out by the UK Cochrane Centre

Copyright © Update Software Ltd.

for all studies relevant to the methodology of systematic reviews. The abstracts presented at all Cochrane Colloquia (1993 to 2001), Systematic Reviews Symposia (1998 to 2002) and Society for Clinical Trials Meetings (1980 to 2001) (as published in Controlled Clinical Trials) have also been handsearched as part of this activity.

The titles and abstracts of records retrieved with these strategies were assessed for relevance to this review (see below, Identifying studies). References in relevant reports were checked to identify additional studies. The Science Citation Index (April 2002) was used to identify articles that cited relevant reports. Finally, researchers who may have carried out relevant studies were contacted. As of June 2002, searches had identified nearly 700 citations and full papers were obtained for 38 records. We made no systematic attempt to identify the use, and influence, of grey literature on the results of individual Cochrane reviews. This will be the subject of a future study.

METHODS OF THE REVIEW

Identifying studies:
One reviewer (SH) screened the titles and abstracts of all retrieved records to identify obvious exclusions (i.e. records that were clearly irrelevant to this review but were found by the electronic searches noted above). A second reviewer (SM) checked less obvious records, before rejection. Full copies of the reports were obtained for each of the non-rejected records. These reports were then assessed independently by at least two reviewers (SH and SM) to determine if they met the inclusion criteria for the review. Any disagreements were resolved through discussion.

Assessment of methodological quality:
The methodological quality of the included studies was assessed against the following criteria: Were explicit criteria used to categorise or define grey literature? Did two or more investigators agree regarding the inclusion of grey literature material? Was there completeness of data for the randomized trials in the meta-analyses included in the empirical study? Any disagreements were resolved through discussion.

Data extraction:
Data extraction was performed independently by two reviewers (SH and SM). Information was extracted for each included empirical study on the area of health care, the number of meta-analyses, the number of randomized trials, the number of participants in these trials, the year of publication of the trials, the language of publication of the trials, the number and type of grey and published literature. Information was also extracted to assess the methodological quality of the included empirical studies. If any of the data for an empirical study were insufficient or missing, these were sought initially from the named contact author (or the responsible person in the case of unpublished studies).

Data analysis:
The decision on whether or not to combine the results of the included empirical studies was dependent on an assessment of heterogeneity. We had hoped to carry out a meta-analysis of those studies which were sufficiently homogenous in their design. However, the summary outcome measures described in the empirical studies were frequently reported differently and therefore could not be combined.

The outcome measures reported by the authors of the empirical studies (for example: relative risk instead of odds ratio) have therefore been presented and analysed descriptively. It is hoped that it may be possible in the future to collect the raw data from each of the empirical studies so that they can be combined in a meta-analysis. It it proves possible to meta-analyse the outcome measures for all of the eight included studies a sensitivity analysis will be performed analysing the single and multiple meta-analyses separately.

Subgroup analysis:
As it was not possible to perform a meta-analysis, data for the following proposed subgroup analyses are presented descriptively:

- the type of grey literature
- the area of health care
- the number of meta-analyses (in particular to distinguish empirical studies of a single meta-analysis from studies of multiple meta-analyses)
- the number of randomized trials
- the number of participants in these trials

DESCRIPTION OF STUDIES

Eight studies met the inclusion criteria for this review (Bhandari 2000; Burdett 2000; Egger 2001; Fergusson 2000; Horn 2002; Jeng 1995; Man-Son-Hing 1998; McAuley 2000). Five of the included studies were published as full papers (Bhandari 2000; Horn 2002; Jeng 1995; Man-Son-Hing 1998; McAuley 2000). Of the remaining three included studies, one study was published as a full paper with additional information obtained from the authors (Fergusson 2000) and two were published as abstracts with additional information obtained from the authors (Burdett 2000; Egger 2001).

Of the eight included studies, four (Burdett 2000; Egger 2001; Fergusson 2000; McAuley 2000) contained multiple meta-analyses, which compared the inclusion and exclusion of grey literature on the pooled effect estimate of the meta-analyses. One of these (Burdett 2000) was also an individual patient data meta-analysis. The remaining four included studies (Bhandari 2000; Horn 2002; Jeng 1995; Man-Son-Hing 1998) were single meta-analyses where the difference between the treatment effect of grey and published trials was explored in a sensitivity analysis. One of these (Man-Son-Hing 1998) was also an individual patient data meta-analysis.

yright © Update Software Ltd.

The included studies covered a variety of different areas of health care. Some studies covered specific areas of health care such as orthopaedic and cardiac surgery, cancer, stroke and miscarriage, while others included meta-analyses from a number of different medical specialities. The meta-analyses included in these studies were identified from a variety of sources, including searches of MEDLINE and specialized trial registers, handsearches of journals and by contacting researchers in the area of interest and the pharmaceutical industry. Within the included studies the definition of what constituted grey literature varied from study to study. However, in general it tended to include abstracts, book chapters, unpublished data, theses, company reports and letters.

More information is available in the Characteristics of included studies table.

METHODOLOGICAL QUALITY

We tried to assess the methodological quality of the included studies using the criteria highlighted above. In all eight included studies explicit criteria were used to categorise or define grey literature. In seven of the included studies two or more investigators agreed regarding the grey literature material included in the meta-analyses. It was not clear if this was done in the study by Jeng 1995. In four of the included studies (Bhandari 2000; Burdett 2000; Egger 2001; McAuley 2000) we judged that there was completeness of data for the included randomized trials. It was not clear if this was the case for the remaining four studies (Fergusson 2000; Horn 2002; Jeng 1995; Man-Son-Hing 1998).

RESULTS

Characteristics of the meta-analyses (Table 01):
Four of the included studies contained multiple meta-analyses and included between 10 and 60 individual meta-analyses. The two studies containing the greatest number of meta-analyses were the studies by Egger 2001 (60 meta-analyses) and McAuley 2000 (41 meta-analyses). Four studies (Bhandari 2000; Horn 2002; Jeng 1995; Man-Son-Hing 1998) contained only one meta-analysis in which a sensitivity analysis had been done to assess the impact of grey literature.

Characteristics of the trials included in the meta-analyses (Table 02):
The number of trials included in the meta-analyses ranged from 7 to 783 trials. Again the studies by Egger 2001 (783 trials) and McAuley 2000 (467 trials) contained the greatest number of trials. In most cases there were more published trials included in the meta-analyses than grey trials (median 46 (IQR 4-300) versus median 5.5 (IQR 4-88). Grey trials are defined as trials in the grey literature, which have not yet been published in full. The number of participants included in the trials ranged from 379 (Jeng 1995) to 217,427 participants (McAuley 2000). Egger 2001 also assessed the statistical significance of the trial results included in the meta-analyses. They found that published trials (30%) were more likely to have statistically significant results ($p<0.05$) than grey trials (19%).

Language of publication of the trials (Table 03):
Information on the language of publication of the trials included in the meta-analyses was available for seven of the eight included studies. The majority of trials (96%) were published in English. This was similar for both grey and published trials and no language restrictions appear to have been imposed by the authors of the studies. Only three of the seven studies (Burdett 2000; Fergusson 2000; McAuley 2000) contained trials published in languages other than English. These included trials published in French, German, Italian, Japanese and Spanish.

Type of grey literature (Table 04):
For all eight included studies the most common type of grey literature were abstracts (49%). Unpublished data was the second largest type of grey literature (33%). The definition of what constituted unpublished data varied since we used the definition used by the authors of the study, however, it included data from trial registers, file-drawer data and data from individual trialists. Book chapters were the third largest type of grey literature (10%), with unpublished reports, pharmaceutical company data, in press publications, letters and theses making up the small remainder. In one study (Man-Son-Hing 1998) pharmaceutical company data was the only source of grey literature. This was an individual patient data meta-analysis where the authors of the study were in direct contact with the company who distributed the drug.

Methodological quality of the randomized trials included in the meta-analyses (Table 05):
Information on the methodological quality of the trials included in the meta-analyses were only available for two of the eight included studies (Bhandari 2000; Egger 2001). Only one of these studies (Egger 2001) contained multiple meta-analyses and they found that the quality of allocation concealment and blinding was better for published trials than for grey trials. In this study the quality of allocation concealment and blinding assessment was only assessed in the subset of trials which were included in Cochrane reviews and it was sometimes unclear or had not been assessed by the authors of Cochrane reviews.

Outcome measures for grey and published trials (Table 06):
The summary outcome measuring the impact of grey and published trials on the pooled effect estimate of the meta-analyses were reported differently across all eight included studies and, therefore, could not be combined formally. Two studies reported a ratio of odds ratios, one an odds ratio, one a relative risk, one a relative risk reduction, one study reported a ratio of hazard ratios and for one study data from the individual trials were not combined.

Multiple meta-analyses
All four of the included studies containing multiple meta-analyses (Burdett 2000; Egger 2001; Fergusson 2000;

Copyright © Update Software Ltd.

McAuley 2000) found that published trials showed an overall greater treatment effect than grey trials. This difference was statistically significant in two of the four studies (Burdett 2000; McAuley 2000). Only one of the included studies (McAuley 2000) assessed the type of grey literature and its impact on the overall results of the meta-analyses. They found that published trials showed an even greater treatment effect than grey trials if abstracts were excluded from the meta-analyses. The ratio of odds ratios for grey trials versus published trials was 1.15 (95% CI 1.04-1.28) compared to 1.33 (95% CI 1.10-1.60) for grey trials, excluding abstracts, versus published trials.

Single meta-analyses

Three of the included studies containing single meta-analyses (Bhandari 2000; Jeng 1995; Man-Son-Hing 1998) found that published trials showed an overall greater treatment effect than grey trials. This difference was not found to be statistically significant. One study, which was a single meta-analysis, found that published trials showed no effect of treatment and that grey trials showed a negative treatment effect (Horn 2002), again this difference was not statistically significant.

DISCUSSION

This review suggests that published trials tend to be larger and may show an overall greater treatment effect than grey trials. On average there were more published trials included in the meta-analyses than grey trials. This may have particular implications in meta-analyses containing only a few trials where the impact of excluding grey trials has the greatest potential to introduce bias.

Only two of the included studies assessed the methodological quality of the trials included in the meta-analyses (the other six did not address this issue). One of these two studies contained multiple meta-analyses (Egger 2001) and found that allocation concealment and blinding were better in published trials than grey trials. However, in this study the methods of allocation concealment and blinding were only assessed in the subset of meta-analyses that were also Cochrane reviews and the necessary information was sometimes unclear or had not been assessed. This highlights possible problems with using the findings of the primary investigators to assess quality and may point to a need for an independent assessment of trial quality.

The definition of grey literature varied across all eight included studies. Overall, abstracts were the most common source of grey literature. One of the included studies (McAuley 2000), which contained data from 41 meta-analyses, found that the exclusion of abstracts from the meta-analyses showed that published trials had an even greater treatment effect than grey trials. This is important as evidence suggests that only around half of all trials reported as abstracts and presented at conferences are subsequently published in full (Scherer 2002). However, there is some controversy as to whether grey literature should be included in meta-analyses as it has not undergone peer review (Cook 1993) and may be of poorer methodological quality (Egger 2001).

REVIEWER'S CONCLUSIONS

Implications for systematic reviews and evaluations of healthcare

This review has important implications for people conducting systematic reviews since they need to ensure they search for trials in the grey, as well as published, literature in order to minimise bias in their review. However, identifying sources of grey trials can be particularly time consuming and involve considerable cost and effort. Within the Cochrane Collaboration substantial efforts are being made to handsearch conference abstracts and other sources of grey literature. Ultimately, however, the prospective registration of all trials is needed in order to identify a complete and unbiased set of trials irrespective of whether they have been published or not.

If and when grey trials are identified their inclusion in a systematic review can sometimes be problematic as the relevant data may often be incomplete or missing and this may be a particular problem in the case of conference abstracts. Those who carry out systematic reviews should therefore contact the authors of these trials for more information. However in absence of more information, data from grey trials is important, particularly when very few trials of a health care intervention have been carried out.

The methodological quality of grey and published trials may also have an impact on the overall results of a systematic review. The CONSORT statement should help improve the quality of reporting of published trials, making the assessment of methodological quality easier for reviewers (Moher 2001). However, the methodological quality of grey trials, in particular conference abstracts, is often poorly reported and difficult to assess. Better reporting of conference abstracts, for example including a description of the methods of allocation concealment and blinding, is needed. In the meantime reviewers should contact the authors of individual studies where the methodological quality of a study is unclear.

Implications for methodological research

A current limitation of this review is that it has not been possible to combine the summary outcome measures from the included studies, as they were frequently reported differently. The issue of how to combine the summary outcome measures is currently being explored and will be addressed in a future version of this review. A meta meta-analysis approach, whereby the data from the individual trials included in the meta-analyses are combined, would help to overcome this problem and also allow greater exploration of issues such as the quality of individual trials and the impact of different types of grey literature on the overall results of the meta-analyse.

Copyright © Update Software Ltd.

Studies which included single meta-analysis were not specifically designed to address the impact of the inclusion and exclusion of grey literature on the overall results of the meta-analysis and had been carried out as sensitivity analyses. These studies are subject to a possible outcome reporting bias as they may only have reported differences between grey and published trials if these were found. If it proves possible to combine and meta-analyse the outcome measures for all of the eight included studies, a sensitivity analysis will be performed analysing the multiple and single meta-analyses separately.

All of the studies included in this review compared binary outcome measures. As far as we are aware no study has assessed the impact of inclusion and exclusion of grey literature in meta-analyses of continuous outcomes.

ACKNOWLEDGEMENTS

We are grateful to Sarah Burdett and Dean Fergusson for supplying unpublished data from their studies.

POTENTIAL CONFLICT OF INTEREST

Sally Hopewell and Mike Clarke work at the UK Cochrane Centre, Oxford and are both involved in the preparation of systematic reviews and aspects of the production of The Cochrane Library. Mike Clarke is also employed by the Clinical Trial Service Unit, Oxford, UK and is involved in the conduct and analysis of randomized trials. Steve McDonald works at the Australasian Cochrane Centre, Melbourne and is involved in methodological research concerning various aspects of systematic reviews. Matthias Egger works at the Department of Social and Preventive Medicine, University of Bern, Switzerland and he is also an author on one of the studies included in this review, which will be published as a NHS Health Technology Assessment review.

SOURCES OF SUPPORT

External sources of support

- No sources of support supplied

Internal sources of support

- UK Cochrane Centre, NHS Research & Development Programme UK
- Australasian Cochrane Centre, Australian Commonwealth Department of Health and Ageing AUSTRALIA
- Department of Social and Preventive Medicine, University of Bern SWITZERLAND

Copyright © Update Software Ltd.

REFERENCES

References to studies included in this review

Bhandari 2000 *{published data only}*
Bhandari M, Guyatt GH, Tong D, Adili A, Shaughnessy SG. Reamed versus nonreamed intramedullary nailing of lower extremity long bone fractures: a systematic overview and meta-analysis. *Journal of Orthopaedic Trauma* 2000;**14**(1):2-9.

Burdett 2000 *{published and unpublished data}*
Burdett S, Stewart L. Publication bias and meta-analysis: a practical example. *8th Annual Cochrane Colloquium Abstracts, Cape Town, October 2000.*

Egger 2001 *{published and unpublished data}*
Egger M, Juni P, Bartlett C, Sterne J. Importance of different sources of bias in systematic reviews of controlled trials: a systematic review of empirical studies. *9th International Cochrane Colloquium, Lyon, October 2001.*

Fergusson 2000 *{published and unpublished data}*
Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPOT meta-analyses. *International Journal of Technology Assessment in Health Care* 2000;**16**(4):1109-1119.

Horn 2002 *{published data only}*
Horn J, Limburg M. Calcium antagonists for acute ischemic stroke. In: *The Cochrane Library*, 1, 2002. Oxford: Update Software.

Jeng 1995 *{published data only}*
Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data. Paternal cell immunization for recurrent miscarriage. *JAMA* 1995;**274**(**10**):830-836.

Man-Son-Hing 1998 *{published data only}*
Man-Son-Hing M, Wells G, Lau A. Quinine for nocturnal leg cramps: a meta-analysis including unpublished data. *Journal of General Internal Medicine* 1998;**13**(**9**):600-606.

McAuley 2000 *{published data only}*
McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analysis. *Lancet* 2000;**356**:1228-1231.

Additional references

Alberani 1990
Alberani V, Pietrangeli P, Mazza A. The use of grey literature in health sciences: a preliminary survey. *Bulletin of the Medical Library Association* 1990;**78**(**4**):358-363.

Auger 1998
Auger CP. *Information sources in grey literature (guides to information sources).* 4th Edition. London: Bowker Saur, 1998.

British Library
British Library, Document Supply Centre. Catalogues and collections - grey literature. *www.bl.uk/services/bsds/dsc/greylit.html*

Cook 1993
Cook D, Guyatt GH, Ryan G. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993;**269**:2749-2753.

Glass 1981
Glass GV, McGaw B, Smith ML. *Meta-analysis in social research.* London: Sage Publication, 1981:65-68.

Loo 1985
Loo JV. Medical and psychological effects of unemployment: a 'grey' literature search. *Health Libraries Review* 1985;**2**:55-62.

Moher 2001
Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;**285**:1992-1995.

Scherer 2002
Scherer RW, Langenberg P. Full publication of results initially presented in abstracts (Cochrane Methodology Review). , The Cochrane Library, Issue 2, 2002. Oxford: Update Software.

SIGLE
SIGLE (System for Information on Grey Literature in Europe). Database summary sheet. *www.cas.org/ONLINE/DBSS/sigless.html [Accessed December 2001]*

Song 2000
Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000;**4**:13-14.

Sterne 2000
Sterne J, Bartlett C, Juni P, Egger M. Do we need comprehensive literature searches? A study of publication and language bias in meta-analyses of controlled trials. *3rd Symposium on Systematic Reviews*. July 2000;Oxford, UK.

Copyright © Update Software Ltd.

TABLES

Characteristics of included studies

Study	Bhandari 2000
Methods	Single meta-analysis carried out as a sensitivity analysis in a systematic review. Trials were identified by searching MEDLINE and SCI Search (1969-1998). Grey literature was defined as abstracts and unpublished manuscripts.
Data	One meta-analysis for reamed versus nonreamed intramedullary nailing of lower long bone fractures, containing 9 RCTs and 714 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.
Outcome measures	Published trials showed a greater treatment effect than grey trials; this difference was not statistically significant. Relative risk 0.43 (95% CI 0.19-0.99) for published trials versus 0.3 (95% CI 0.11-0.83) for grey trials.
Notes	

Study	Burdett 2000
Methods	Multiple individual patient data meta-analyses. Trials were identified by electronic searches, handsearching, trial registers and by contacting trialists. Grey literature was defined as unpublished data, abstracts, book chapters and non-English language.
Data	11 meta-analyses of cancer treatments, containing 120 RCTs and 18,377 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.
Outcome measures	Published trials showed a greater treatment effect than grey trials; this difference was statistically significant. Ratio of hazard ratios 0.93 (95% CI 0.90-0.97) for published trials versus 0.96 (95% CI 0.93-0.99) for all trials.
Notes	

Study	Egger 2001
Methods	Multiple meta-analyses were identified by handsearching 8 high-impact medical journals (1994-1998), searching CDSR (issue 1, 1998), DARE (1994-1998) and handsearching HTA reports. Meta-analyses were eligible if they combined binary outcomes of at least 5 trials and had comprehensive literature searches. Grey literature was defined as conference abstracts, books, theses, file drawer data and material from trial registers.
Data	60 meta-analyses from a number of different medical specialties, containing 783 RCTs and 167,733 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.
Outcome measures	Published trials showed a greater treatment effect; this difference was not statistically significant. Ratio of odds ratio 1.07 (95% CI 0.98-1.15) for grey trials versus published trials.
Notes	

Copyright © Update Software Ltd.

Characteristics of included studies

Study	Fergusson 2000
Methods	Multiple meta-analyses identified from a project undertaken by the International Study of Peri-operative Transfusion. Trials were identified by searching MEDLINE and EMBASE (1966-1996). Grey literature was defined as abstracts, letters and conference proceedings.
Data	10 meta-analyses of methods to reduce the need for peri-operative transfusion, containing 114 trials and 11,142 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.
Outcome measures	Published trials showed a greater treatment effect than grey trials; this difference was not statistically significant.
Notes	

Study	Horn 2002
Methods	Single meta-analysis carried out as a sensitivity analysis in a Cochrane review. Trials were identified by searching the Cochrane Stroke Review Group's specialised register. Grey literature was defined as abstracts and unpublished data.
Data	One meta-analysis for calcium antagonists for acute ischaemic stroke, containing 21 RCTs and 6414 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.
Outcome measures	Published trials showed no treatment effect. Grey trials showed a negative treatment effect. This difference was not statistically significant. Odds ratio 1.03 (95% CI 0.92-1.15) for published trials versus 1.33 (95% CI 1.00-1.79) for grey trials.
Notes	

Study	Jeng 1995
Methods	Single meta-analysis carried out as a sensitivity analysis in an individual patient data meta-analysis. Trials were identified through the ethics committee of the American Society for Reproductive Immunology. Grey literature was defined as unpublished.
Data	One meta-analysis of paternal immunisation for recurrent miscarriage, containing 8 trials and 379 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.
Outcome measures	Published trials showed a greater treatment effect; this difference was not statistically significant. Relative risk 1.17 (95% CI 0.97-1.37) for published trials versus 1.01 (95% CI 0.74-1.28) for grey trials.
Notes	

Study	Man-Son-Hing 1998
Methods	Single meta-analysis carried out as a sensitivity analysis in a systematic review using individual patient data. Trials were identified by searching MEDLINE and EMBASE (1994-1997). Grey literature was defined as data obtained from the pharmaceutical industry.

Copyright © Update Software Ltd.

Characteristics of included studies

Data	One meta-analysis of quinine for nocturnal leg cramps, containing 7 trials and 409 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analysis.
Outcome measures	Published trials showed a greater treatment effect; this difference was not statistically significant. Relative risk reduction 0.43 (95% CI 0.21-0.65) for published trials versus 0.15 (95% CI 0.06-0.24) for grey trials.
Notes	

Study	**McAuley 2000**
Methods	Multiple meta-analyses were selected at random from an existing database of meta-analyses established by searching MEDLINE (1966-1995). Grey literature was defined as abstracts, unpublished studies, conference proceedings, theses, book chapters and company reports.
Data	41 meta-analyses from a number of different medical specialties, containing 467 RCTs and 217,427 participants.
Comparisons	Inclusion and exclusion of grey literature on pooled effect estimate of the meta-analyses.
Outcome measures	Published trials showed a greater treatment effect; this difference was statistically significant. Ratio of odds ratio 1.15 (95 % CI 1.04-1.28) for grey versus published trials. Ratio of odds ratio 1.02 (95% CI 0.91-1.14) for abstracts versus published trials. Ratio of odds ratio 1.33 (95 % CI 1.10-1.60) for grey (excluding abstracts) versus published trials.
Notes	

ADDITIONAL TABLES

Table 01 Characteristics of the meta-analyses

Study	No. meta-analyses	No. trials	No. participants	No. grey trials	No. pub trials	Area healthcare	Year publication
Bhandari 2000	1	9	714	5	4	Orthopaedic surgery	1996-1998
Burdett 2000	11	120	18,377	45	75	Cancer	
Egger 2001	60	783	167,733	153	630	Various medical specialties	1994-1998
Fergusson 2000	10	114	11,142	6	108	Cardiac and Orthopaedic surgery	
Horn 2002	1	21	6,414	4	17	Acute ischemic stroke	1984-1999

Copyright © Update Software Ltd.

Table 01 Characteristics of the meta-analyses

Jeng 1995	1	8	379	4	4	Paternal immunisation for recurrent miscarriage	1986-1993
Man-Song-Hing 1998	1	7	409	3	4	Quinine for nocturnal leg cramps	1987-1993
McAuley 2000	41	467	217,427	102	365	Gastrointestinal (24%), Cardiac (21%), Infection (12%), Reproduction (12%), Circulatory (9%), Other (21%)	

Table 02 Characteristics of the trials included in meta-analyses

Study	No. trials	No. participants	No. per trial	Positive trials	Negative trials
Bhandari 2000	Grey 5; Published 4	Grey 348; Published 366	Grey (median) 75 (IQR 49-87.5); Published (median) 87.5 (IQR 57.7-129.2)		
Burdett 2000	Grey 45; Published 75	Grey 6,221; Published 12,156	Grey (mean) 138; Published (mean) 162		
Egger 2001	Grey 153; Published 630	Grey 21,573; Published 146,160	Grey (median) 91 (range 9-1012); Published (median) 102 (range 8-5042)	p<0.05: Grey 29 (19%); Published 187 (30%)	p<0.01: Grey 18 (12%); Published 100 (16%)
Fergusson 2000	Grey 6; Published 108	Grey 820; Published 10,322	Grey (median) 149 (IQR 85-250); Published (median) 638 (IQR 240-828)		
Horn 2002	Grey 4; Published 17	Grey 788; Published 5,626	Grey (median) 219 (IQR 135-237); Published (median) 189 (IQR 55-411)		

Copyright © Update Software Ltd.

Table 02 Characteristics of the trials included in meta-analyses

Jeng 1995	Grey 4; Published 4	Grey 140; Published 239	Grey (median) 33 (IQR 24-48); Published (median) 54 (IQR 40-85)		
Man-Song-Hing 1998	Grey 3; Published 4	Grey 336; Published 73	Grey (median) 69 (IQR 62-205); Published (median) 19 (IQR 10-26)		
McAuley 2000	Grey 102; Published 365	Grey 23,286; Published 194,141	Grey (median) 84 (IQR 48-190); Published (median) 113 (IQR 59-228)	p<0.05: Grey 70 (68%); Published 240 (66%)	p>=0.05: Grey 32 (32%); Published 125 (35%)

Table 03 Language of publication of the trials

Study	English	French	German	Spanish	Italian	Other
Bhandari 2000	Grey 5 (100%); Published 4 (100%)					
Burdett 2000	Grey 44 (98%); Published 75 (100%)					Grey 1 (2%)
Fergusson 2000	Overall 100 (88%)	Overall 6 (4%)	Overall 4 (3%)	Overall 2 (2%)	Overall 3 (2%)	Overall 1 Belgian (1%)
Horn 2002	Grey 4 (100%); Published 17 (100%)					
Jeng 1995	Grey 4 (100%); Published 4 (100%)					
Man-Song-Hing 1998	Grey 3 (100%); Published 4 (100%)					
McAuley 2000	Grey 99 (97%); Published 340 (93%)	Grey 2 (2%); Published 9 (2%)	Published 9 (2%)	Grey 1 (1%); Published 3 (1%)	Published 3 (1%)	Japanese: Published 1

Table 04 Type of grey literature

Study	Abstract	Unpublished	Book chapter	Report	Drug company	In press	Letter	Thesis	Other
Bhandari 2000	4 (80%)	1 (20%)							

Copyright © Update Software Ltd.

Table 04 Type of grey literature

Burdett 2000	17 (38%)	24 (53%)	3 (7%)						1 (2%) Non-English language
Egger 2001	69 (45%)	57 (37%)	22 (14%)					5 (3%)	
Fergusson 2000	4 (67%)	1 (17%)			1 (17%)				
Horn 2002	2 (50%)	2 (50%)							
Jeng 1995		4 (100%)							
~~Man~~					3 (100%)				
McAuley 2000	63 (62%)	18 (17%)	6 (6%)	5 (5%)	3 (3%)	3 (3%)	2 (2%)	2 (2%)	

Table 05 Methodological quality of the randomized trials included in the meta-analyses

Study	No. trials	Concealment YES	Concealment NO	No. trials	Blinding YES	Blinding NO
Bhandari 2000	Grey 5; Published 4	Grey 2 (40%); Published 3 (75%)	Grey 3 (60%); Published 1 (25%)	N/A		
Egger 2001	Grey 77; Published 339	Grey 26 (34%); Published 138 (41%)	Grey 51 (66%); Published 201 (59%)	Grey 71; Published 345	Grey 32 (45%); Published 227 (66%)	Grey 39 (55%); Published 118 (34%)

Table 06 Summary outcome measures for grey and published trials

Study	Overall	Grey	Published	Conclusion
Bhandari 2000	Relative risk 0.33 (95% CI 0.16-0.68)	Relative risk 0.30 (95% CI 0.11-0.83)	Relative risk 0.43 (95% CI 0.19-0.99)	Published trials showed a greater treatment effect in favour of the experimental intervention; this was not statistically significant.
Burdett 2000	Ratio of hazard ratios 0.96 (95% CI 0.93-0.99)		Ratio of hazard ratios 0.93 (95% CI 0.90-0.97)	Published trials showed a greater treatment effect in favour of the experimental intervention; this was statistically significant.

Copyright © Update Software Ltd.

Table 06 Summary outcome measures for grey and published trials

Egger 2001	Ratio of odds ratios (grey versus published) 1.07 (95% CI 0.98-1.15)			Published trials showed a greater treatment effect in favour of the experimental intervention; this was not statistically significant.
Fergusson 2000				Published trials showed a greater treatment effect in favour of the experimental intervention; this was not statistically significant.
Horn 2002	Odds ratio 1.06 (95% CI 0.96-1.18)	Odds ratio 1.33 (95% CI 1.00-1.79)	Odds ratio 1.03 (95% CI 0.92-1.15)	Published trials showed no effect of treatment. Grey trials showed a negative treatment effect; this was not statistically significant.
Jeng 1995		Relative risk 1.01 (95 % CI 0.74-1.28)	Relative risk 1.17 (95% CI 0.97-1.37)	Published trials showed a greater treatment effect in favour of the experimental intervention; this was not statistically significant.
Man-Song-Hing 1998	Relative risk reduction 0.21 (95% CI 0.12-0.30)	Relative risk reduction 0.15 (95% CI 0.06-0.24)	Relative risk reduction 0.43 (95% CI 0.21-0.65)	Published trials showed a greater treatment effect in favour of the experimental intervention; this was not statistically significant.
McAuley 2000	Ratio of odds ratio (grey versus published) 1.15 (95% CI 1.04-1.28)			Published trials showed a greater treatment effect in favour of the experimental intervention; this was statistically significant.

Copyright © Update Software Ltd.

APPENDIX 2: Analysis of estimates of treatment effect

This appendix contains data from the analysis of estimates of treatment effects for published and grey trials described in Chapter Three of this thesis.

Analysis of estimates of treatment effects for published and grey trials (Chapter Three)

Review	n_overall	n_published	n_grey	n_personal	or_overall	selogor_overall	or_published	selogor_published	or_grey
Review 1	10	6	1	3	0.949	0.102	1.155	0.118	0.319
Review 2	9	8	1	0	0.708	0.149	0.711	0.152	0.643
Review 3	8	1	5	2	7.919	0.416	69.841	1.427	3.275
Review 4	4	0	3	1	5.852	0.763			7.095
Review 5	8	6	0	2	0.987	0.016	0.985	0.017	
Review 6	15	13	2	0	0.744	0.065	0.727	0.072	0.834
Review 7	7	5	0	2	0.766	0.174	0.768	0.183	
Review 8	14	8	0	6	0.878	0.118	0.899	0.144	
Review 9	26	17	4	5	0.829	0.098	0.831	0.136	0.653
Review 10	11	9	2	0	0.631	0.082	0.573	0.090	1.024
Review 11	8	6	2	0	0.209	0.171	0.205	0.180	0.239
Review 12	22	13	2	7	0.400	0.085	0.387	0.097	0.980
Review 13	2	1	1	0	0.091	0.159	0.063	0.256	0.114
Review 14	10	9	1	0	0.982	0.094	0.971	0.100	1.087
Review 15	11	1	1	9	0.874	0.121	0.872	0.316	0.250
Review 16	9	5	0	4	0.494	0.237	0.504	0.307	
Review 17	3	2	1	0	0.506	0.458	0.781	0.667	0.341
Review 18	10	9	1	0	0.157	0.309	0.194	0.322	0.020
Review 19	14	13	1	0	0.878	0.058	0.874	0.060	0.923
Review 20	14	11	0	3	0.826	0.102	0.727	0.121	
Review 21	16	10	6	0	0.507	0.086	0.518	0.120	0.496
Review 22	6	5	1	0	0.523	0.194	0.493	0.211	0.746
Review 23	51	49	2	0	0.603	0.046	0.605	0.047	0.547
Review 24	10	8	1	1	0.732	0.097	0.759	0.102	0.442
Review 25	8	7	1	0	0.693	0.147	0.790	0.173	0.493

n = number of trials
or = odds ratio
selogor = standard error of log odds ratio

Analysis of estimates of treatment effects for published and grey trials (continued)

Review	selogor_grey	or_personal	selogor_personal	het_overall	het_published	het_grey	het_personal	wt_published	wt_grey	wt_personal
Review 1	0.352	0.712	0.262	26.449	9.049	0.000	4.194	67.168	15.431	17.401
Review 2	0.765			24.207	24.199	0.000		95.847	4.153	0.000
Review 3	0.541	2.375	0.990	9.223	0.000	2.183	0.034	7.283	70.059	22.658
Review 4	0.874	2.493	1.640	1.081		0.871	0.000	0.000	72.995	27.005
Review 5		1.014	0.070	8.980	8.810		0.003	94.702	0.000	5.298
Review 6	0.157			58.251	49.363	8.375		83.680	16.320	0.000
Review 7		0.752	0.567	2.370	2.201		0.169	90.333	0.000	9.667
Review 8		0.837	0.206	5.210	3.686		1.422	66.416	0.000	33.584
Review 9	0.290	0.892	0.164	21.498	12.019	5.276	3.667	52.332	13.033	34.635
Review 10	0.201			20.010	12.253	0.728		87.112	12.888	0.000
Review 11	0.528			31.434	29.847	1.440		89.254	10.746	0.000
Review 12	0.493	0.400	0.192	74.334	55.167	1.404	14.322	78.477	1.874	19.650
Review 13	0.203			3.364	0.000	0.000		46.218	53.782	0.000
Review 14	0.289			14.134	13.994	0.000		89.857	10.143	0.000
Review 15	1.204	0.889	0.132	8.095	0.000	0.000	6.985	14.651	2.098	83.251
Review 16		0.479	0.373	13.810	3.776		10.073	58.670	0.000	41.330
Review 17	0.642			2.320	1.587	0.000		37.559	62.441	0.000
Review 18	1.458			11.435	8.329	0.000		78.396	21.604	0.000
Review 19	0.199			16.782	16.709	0.000		91.844	8.156	0.000
Review 20		1.139	0.192	25.995	22.370		0.338	76.012	0.000	23.988
Review 21	0.123			23.689	9.562	14.096		49.312	50.688	0.000
Review 22	0.512			4.203	3.701	0.000		88.197	11.803	0.000
Review 23	0.269			62.035	60.053	1.845		96.877	3.123	0.000
Review 24	0.763	0.530	0.363	16.021	14.625	0.000	0.000	89.240	2.153	8.606
Review 25	0.281			8.650	6.626	0.000		67.377	32.623	0.000

het = test for heterogeneity
wt = weight contributed in the meta-analysis

APPENDIX 3: Data extraction form

This appendix contains the data extraction form for the study described in Chapter Six of this thesis comparing the information reported for a trial in a conference abstract with that reported in its subsequent full publication.

Comparing abstracts with their subsequent full publication

Abstract number:
Reviewer name:

Year and month of publication of abstract:
Year and month of publication of paper:

Authors (individual or group authorship):

Authors (describe any changes to the authors between the abstract and full paper)**:**

Title (describe any changes to the title between the abstract and full paper)**:**

Multicentre status (multi / single / not stated):

Length of published report (no. pages)**:**

Funding source:

Ethical approval:

Criteria	Item	Descriptor	Abstract	Paper	Type of change
Authors	1	Adequate correspondence details			
Objectives	2a	Clearly state study objectives			
	2b	Trial name or identifier given			
	2c	Date of trial given (i.e. any date which fixes the trial in time)			
Study design	3a	"Trial" in the title			
	3b	"Randomized" in the title			
Study quality	4a	Method of allocation concealment described adequate inadequate not stated			
	4b	Method of blinding described adequate inadequate not stated			

Criteria	Item	Descriptor	Abstract	Paper	Type of change
Participants	5	Broad characteristics of eligible participants described (i.e. inclusion and/or exclusion criteria)			
Interventions	6a	Experimental intervention described			
	6b	What was the experimental intervention?			
	6c	Comparator intervention(s) described			
	6d	What was the comparator intervention(s)?			
	6e	Number of participants randomized (eligible) to experimental intervention			
	6f	Number of participants randomized (eligible) to comparator intervention(s)			
Outcomes	7a	What was the main outcome assessed?			
	7b	Does the abstract / paper describe the trial as closed yes ongoing not clear			
	7c	Total number of participants randomized			
	7d	Number of participants analysed			
	7e	Intention to treat analysis stated or reason for exclusion given			
	7f	Important adverse / side effects reported in each intervention group			
Results	8a	Does the paper present the same results as the abstract?			
	8b	Summary result for each group and the estimated effect size and its precision (e.g. 95% CI) (if not, what was reported?)			
Conclusions	9	Primary conclusion(s) described			

APPENDIX 4: Abstract comparing trials reported as abstracts and full papers

The checklist and methods developed as part of this thesis (Chapter Six) have subsequently been used in a study of surgical trials in Australia. This appendix contains the abstract for that study, which compares the information reported for a trial in a conference abstract with that reported in its subsequent full publication:

- Tooher, R., Middleton, P., Griffin, T., Pham, C. and Hopewell, S. (2004) How different are conference abstracts of surgical RCTs from their subsequent full publication? *In: 12th International Cochrane Colloquium, 2003 Oct 2-6, Ottawa, Canada* (forthcoming).

How different are conference abstracts of surgical RCTs from their subsequent full publication?

Tooher R, Middleton P, Griffin T, Pham C, Hopewell S.

Background: Systematic reviewers of new surgical procedures are often forced to rely on low level and/or poor quality evidence because randomised controlled trials (RCTs) of new surgical procedures are relatively rare. As a result, reviewers may be faced with the decision of whether to include conference abstracts of surgical RCTs in systematic reviews or only include evidence which has been fully published, including lower level evidence.

Objectives: This study compared conference abstracts of surgical RCTs to their subsequent full publication to determine the level of agreement between the publications.

Methods: The conference proceedings of three surgical specialty conferences (urology, orthopaedics, gastroendoscopy) were handsearched for abstracts of randomised controlled trials. Only trials of surgical procedures on human subjects were included. The Cochrane Central Register of Controlled Trials, MEDLINE and EMBASE (1998-Nov 2003) were searched to identify subsequent full publications. Data regarding study design, characteristics and numbers of participants, outcomes reported and conclusions were extracted by two reviewers and checked by a third. Any disagreements were resolved by discussion. The level of concordance between abstracts and full publications was summarised into a percentage score for key outcomes.

Results: Of 1843 abstracts, only 92 were reports of RCTs and of these only 45 (2.4%) were surgical. 39/45 (87%) conference abstracts had been published in full in the five years since presentation at a conference (average time to publication 22.3 months, range 2 – 66). Two non-English language studies were excluded leaving 37 abstracts and full publications for analysis. The quality of reporting in abstracts was poorer than in full publications, due, at least in part, to space restrictions; however, some aspects of reporting were poor in both abstracts and full publications. For example, identifying details were broadly similar, with 68% of first authors and 62% of study titles the same in both abstract and full publication. Method of allocation concealment was only reported in 16 (43%) full publications and was not reported in any abstract. Method of blinding was not reported in 31 (84%) abstracts and 23 (62%) full publications. More participants were randomised in full publications (median 81) than in abstracts (median 60) with 7 abstracts reporting interim results. In terms of participants analysed, only one abstract and 3 full publications clearly stated use of intention-to-treat, with 2 abstracts and 19 full publications providing reasons for losses to follow-up. As a result, fewer participants were analysed in 8 full publications as no losses to follow-up were assumed in abstracts unless stated. Overall, 24% of abstracts and full publications reported the same numbers of participants randomised and analysed. Due to these differences in numbers of participants included, results reported were only the same for 45% (15/33) of abstracts and full publications; however, the direction of results was the same in 70% (23/33) (Concordance could not be determined in four studies due to lack of detail in abstracts). The primary conclusion was the same in 32/37 (87%) studies.

Conclusions: While it is difficult to judge the methodological quality of abstracts, this problem also exists in many full reports of surgical RCTs. As the direction of results and the conclusions drawn rarely differed, abstracts of surgical RCTs can be included in systematic reviews, but caution must be exercised. Reviewers should attempt to contact authors for additional study details, and abstracts included in meta-analysis should be clearly identified and sensitivity analysis performed to allow readers to judge the validity of the findings.