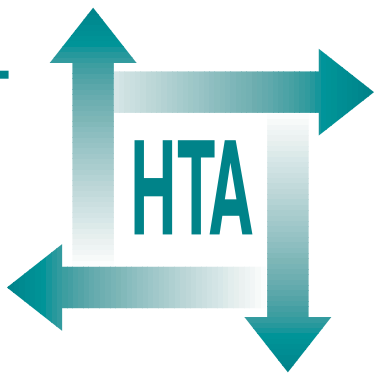
Health Technology Assessment 2000; Vol. 4: No. 10

Methodology

Publication and related biases

F Song
AJ Eastwood
S Gilbody
L Duley
AJ Sutton

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Current members

Chair:
Professor Kent Woods
Professor of Therapeutics,
University of Leicester

Professor Martin Buxton
Director & Professor of
Health Economics,
Health Economics
Research Group,
Brunel University

Professor Shah Ebrahim
Professor of Epidemiology
of Ageing, University of Bristol

Professor Francis H Creed
Professor of
Psychological Medicine,
Manchester Royal Infirmary

Professor John Gabbay
Director, Wessex Institute
for Health Research
& Development

Professor Sir John
Grimley Evans
Professor of Clinical Geratology,
Radcliffe Infirmary,
Oxford

Dr Tony Hope
Clinical Reader in Medicine,
Nuffield Department of
Clinical Medicine,
University of Oxford

Professor Richard Lilford
Regional Director of R&D,
NHS Executive West Midlands

Dr Jeremy Metters
Deputy Chief Medical Officer,
Department of Health

Professor Maggie Pearson
Regional Director of R&D,
NHS Executive North West

Mr Hugh Ross
Chief Executive,
The United Bristol
Healthcare NHS Trust

Professor Trevor Sheldon
Joint Director, York Health
Policy Group, University of York

Professor Mike Smith
Faculty Dean of Research
for Medicine, Dentistry,
Psychology & Health,
University of Leeds

Dr John Tripp
Senior Lecturer in Child
Health, Royal Devon and Exeter
Healthcare NHS Trust

Professor Tom Walley
Director,
Prescribing Research Group,
University of Liverpool

Dr Julie Woodin
Chief Executive,
Nottingham Health Authority

Past members

Professor Sir Miles Irving*
Professor of Surgery,
University of Manchester,
Hope Hospital, Salford

Dr Sheila Adam
Department of Health

Professor Angela Coulter
Director, King's Fund,
London

Professor Anthony Culyer
Deputy Vice-Chancellor,
University of York

Dr Peter Doyle
Executive Director, Zeneca Ltd,
ACOST Committee on Medical
Research & Health

Professor John Farndon
Professor of Surgery,
University of Bristol

Professor Charles Florey
Department of Epidemiology
& Public Health, Ninewells
Hospital & Medical School,
University of Dundee

Professor Howard
Glennester
Professor of Social Science
& Administration, London
School of Economics &
Political Science

Mr John H James
Chief Executive,
Kensington, Chelsea &
Westminster Health Authority

Professor Michael Maisey
Professor of
Radiological Sciences,
Guy's, King's & St Thomas's
School of Medicine & Dentistry,
London

Mrs Gloria Oates
Chief Executive,
Oldham NHS Trust

Dr George Poste
Chief Science & Technology
Officer, SmithKline Beecham

Professor Michael Rawlins
Wolfson Unit of
Clinical Pharmacology,
University of Newcastle-
upon-Tyne

Professor Martin Roland
Professor of General Practice,
University of Manchester

Professor Ian Russell
Department of Health Sciences
& Clinical Evaluation,
University of York

Dr Charles Swan
Consultant Gastroenterologist,
North Staffordshire
Royal Infirmary

* Previous Chair

Details of the membership of the HTA panels, the NCCHTA Advisory Group and the HTA Commissioning Board are given at the end of this report.

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Publication and related biases

F Song[1*]
AJ Eastwood[1]
S Gilbody[1]
L Duley[2]
AJ Sutton[3]

[1] NHS Centre for Reviews and Dissemination, University of York, York, UK
[2] Magpie Trial, Institute of Health Sciences, Oxford, UK
[3] Department of Epidemiology and Public Health, University of Leicester, Leicester,UK

[*] Corresponding author

Competing interests: none declared

Published July 2000

This report should be referenced as follows:

Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;**4**(10).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see overleaf).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Group and funded as project number 95/12/01.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work either prioritised by the Standing Group on Health Technology, or otherwise commissioned for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

Contents

List of abbreviations i

Executive summary iii

1 Background 1
Definition of publication and related biases 1

2 Review methods 5
Systematic review of relevant literature 5
Survey of published systematic reviews 5

3 Research findings and dissemination profiles 7
Indirect evidence 7
Direct evidence 9
Other publication-related biases 13
Summary 20

4 Consequences of publication and related biases 23
Patients' suffering and waste of resources ... 24
Controversies due to publication and related biases 24
Summary 25

5 Sources of publication bias 27
Who is responsible for publication bias? 27
Statistical considerations 32
Summary 34

6 Prevention of publication bias 35
Changes in editorial policy and the publication process 35
Prospective registration of trials 37
Freedom of information 39
Confirmatory large-scale trials 39
Summary 40

7 Reducing or detecting publication and related biases in systematic reviews 41
Literature searching 41
Locating unpublished trials 41
Assessing the possibility of publication bias 42
Funnel plot 42
Statistical and modelling methods 45
Fixed or random effects models 50
Updating systematic reviews 50
Summary 50

8 Survey of published systematic reviews 53
Literature searching in the reviews 53
Unpublished studies 54
Non-English language studies 54
Use of methods for publication bias 55
Evidence provided on publication bias 55
Conclusions of surveyed systematic reviews 55
Summary 55

9 Discussion and recommendations 59
Publication ethics 59
Publication bias and other sources of bias 60
Implications for decision makers and researchers 60
Recommendations for future research 61

Acknowledgements 63

References 65

Appendix 1 Strategies for searching electronic databases 77

Appendix 2 Data extraction sheet for methodological/empirical studies 79

Appendix 3 Data extraction sheet for surveying DARE systematic reviews 81

Appendix 4 Stochastic simulation for publication bias 83

Appendix 5 A list of registries of clinical trials 85

Appendix 6 Modelling for publication bias in meta-analysis 87

Health Technology Assessment reports published to date 107

Health Technology Assessment panel membership 111

List of abbreviations

BMJ	*British Medical Journal*
CCTR	Cochrane Controlled Trials Register
CI	confidence interval
DARE	Database of Abstracts of Reviews of Effectiveness
ETS	environmental tobacco smoking
FDA	United States Food and Drug Administration
IRB	institutional review board
JAMA	*Journal of the American Medical Association*
lnOR	$\log_e$ (odds ratio)
NEJM	*New England Journal of Medicine*
NIH	National Institutes of Health
NNT	number needed to treat
NSAID	non-steroidal anti-inflammatory drug
OR	odds ratio
RCT	randomised controlled trial
R&D	research and development
REC	research ethics committee
RR	relative risk, or rate ratio
SE	standard error
SND	standard normal deviate
TENS	transcutaneous electrical nerve stimulation

Executive summary

Background

Literature review is becoming increasingly important in summarising research evidence for clinical and health policy decision making because of a rapidly expanding volume of medical research. However, the results of literature reviews will be misleading if the published studies comprise a biased sample of all the studies that have been conducted.

The term "dissemination profile" could be used to describe the accessibility of research results or the possibility of research findings being identified by potential users. The spectrum of the dissemination profile ranges from completely inaccessible to easily accessible, according to whether, when, where and how research is published.

Objectives

This review aimed to identify systematically and appraise studies that have examined methodological issues and provided empirical evidence about publication bias and other dissemination-related biases, including biases due to the time, type and language of publication, multiple publication, selective citation of references, database index bias, and biased media attention. The review sought to answer the following questions:

- What empirical evidence is available on the existence and consequences of publication and related biases?
- What are the causes and risk factors of publication and related biases?
- What methods have been developed and how useful are these methods for preventing, detecting and correcting publication and related biases?

Methods

This report includes a systematic review of publication and related biases, and a survey of publication bias in published systematic reviews.

Systematic review of publication and related biases

The following databases were searched to identify relevant literature concerning empirical evidence and methodological issues pertaining to publication and related biases: the Cochrane Review Methodology Database, MEDLINE, EMBASE, BIDS, Library and Information Science Abstracts, PsycINFO, Sociofile, ERIC, Dissertation Abstracts, MathSci, British Education Index, SIGLE and ASSIA. The reference lists of the identified articles were also checked.

The results of searches of electronic databases were checked independently by two reviewers and any disagreements discussed. Full publications for studies that were considered to be potentially relevant were obtained and their suitability for inclusion independently assessed by at least two reviewers. All studies relevant to publication and related biases were included, except if the issue of publication bias was not a major topic. Data from included studies were collected by one reviewer by using a data-extraction form and then checked by another reviewer.

Survey of published systematic reviews

A sample of 193 systematic reviews was taken from the Database of Abstracts of Reviews of Effectiveness (NHS Centre for Reviews and Dissemination at the University of York) to identify further evidence of publication and related biases and to illustrate the methods used for dealing with publication bias. These reviews were assessed independently by two reviewers using a data-extraction form.

Results

Research findings and dissemination profiles

The empirical evidence demonstrates that studies with significant results or favourable results are more likely to be published or cited than those with non-significant or unfavourable results. Studies with significant results are often published earlier than those with non-significant results. Limited and often indirect evidence indicates only the possibility of full publication bias,

outcome reporting bias, duplicate publication bias, language bias and database bias. There is some evidence concerning the existence of citation bias and media attention bias.

Consequences of publication and related biases

The important consequences of publication bias include the avoidable suffering of patients and the waste of limited resources. However, there is little empirical evidence relating to the impact of publication and related biases on health policy, clinical decision making and the outcome of patient management.

Sources of publication bias

Investigators, peer reviewers, editors and funding bodies may all be responsible for the existence of publication bias. Some evidence suggests that authors or investigators may be the main source of this bias, for not writing up or not submitting studies with null or unimportant results. However, it should be recognised that the decision to write up an article and then submit it may be affected by pressure from research sponsors and instruction from journal editors. Evidence shows that the interest of research sponsors can restrict the dissemination of research findings. The large potential variation in results obtained across similar studies that can easily be conducted and abandoned will further exacerbate the biased selection of findings for publication.

Prevention of publication bias

Because of their space limitations and need to maintain newsworthiness, it is unlikely that conventional paper journals can solve the problem of the selective publication of studies that produce striking results. For the purpose of reducing publication bias, peer-reviewed electronic journals that are without limitations of space are required. More importantly, editorial policy needs to be changed to accept for publication clinical trials that are based on methodological criteria only and not on the impact of their findings.

Clearly, the ideal solution to publication bias is the prospective, universal registration of all studies at their inception. Although the registration of all studies cannot be realised in the near future, there are many encouraging signs that there will be more registries established as a result of initiatives from government or industry. Large-scale confirmatory studies may be an alternative in the prevention of the consequences of publication bias.

Methods for reducing or detecting publication bias

The methods available for dealing with publication and related biases in systematic reviews include literature searching, locating unpublished studies, assessment of the risk of publication and related biases, several methods for detecting publication bias in meta-analyses, and updating systematic reviews. The statistical methods are by nature indirect and exploratory, and often based on certain strict assumptions that can be difficult to justify in the real world. The attempt at identifying or adjusting for publication bias in a systematic review should mainly be used for the purpose of sensitivity analysis.

Survey of published systematic reviews

This survey indicates that literature searching was clearly inadequate in some published systematic reviews. Potential publication bias was ignored and the available methods for dealing with such bias were not used in most of these reviews. When they are used to estimate possible publication bias at the stage of literature review, the available methods were far from adequate and their usefulness was strictly limited. The problem of publication and related biases was dealt with more often in reviews containing a meta-analysis than in the narrative systematic reviews.

Conclusions

Although the extent, direction and impact of publication and related biases are uncertain and may vary greatly depending on circumstances, it seems reasonable to conclude that studies with significant or favourable results are more widely disseminated than those with non-significant or unfavourable results. The potential problem of publication and related biases should be taken into consideration in the field of health technology assessment. All funded or approved studies should be prospectively registered. The risk of publication bias should be assessed in all systematic reviews.

Recommendations for future research

- Further research is needed to provide more direct empirical evidence about publication and related biases. In particular, there is a lack of evidence about the impact of publication bias on health decision making and the outcomes of patient management.

- The available methods for dealing with publication bias should be evaluated by comparing their assumptions, performance and results, ideally by using a set of meta-analyses in which the extent of publication bias could be estimated according to unbiased samples of relevant studies.
- Research is also needed to develop new methods that are robust and easy to use for detecting publication bias in systematic reviews. In particular, there is a lack of methods that can be used to detect publication bias in narrative systematic reviews.
- Further research is needed to answer questions about: how to establish and maintain the prospective registration of clinical trials and observational studies; how to make all research findings accessible to the public; and how the developments in computer science and information technology can be used to solve the problem of publication bias.
- Further research concerning publication bias should be an integral part of research that explores alternatives to the conventional methods for generating, disseminating, preserving and utilising scientific research findings.

Chapter 1
Background

Literature review is becoming increasingly important in summarising research evidence for clinical and health policy decision making. A key factor in this growing reliance on the results of reviews is the rapidly expanding volume of medical research.[1] In contrast to traditional narrative reviews, which have been criticised for being subjective, scientifically unsound and inefficient,[2] systematic reviews produce more reliable results by methodically locating, appraising and synthesising research evidence.[3]

Both traditional and systematic reviews are heavily dependent on the published literature. However, the number of studies included in a literature review may not be equal to the number of all relevant studies conducted because of either a failure to publish or difficulties in locating some publications (*Figure 1*). If studies missing from a review have results that are systematically different from included studies, biases or systematic errors will occur.[4]

The observation that many studies are never published has been termed "the file-drawer problem"[5] or "the iceberg phenomenon".[6] The importance of this problem depends on whether or not the published studies are representative of all studies that have been conducted. If the published studies are the same as, or a random sample of, all studies that have been conducted, there will be no bias and the average estimate based on the published studies will be similar to that based on all studies. If the published studies comprise a biased sample of all studies that have been conducted, the results of a literature review will be misleading.[7] For example, the efficacy of a treatment will be overestimated if studies with positive results are more likely to be published than those with negative results.

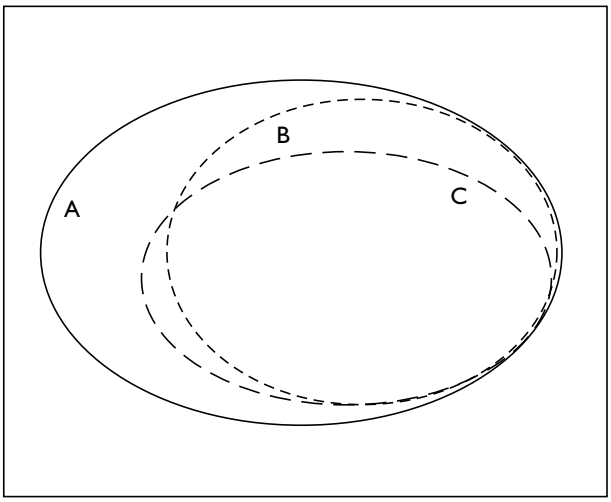


FIGURE 1 *Many completed studies may not be published and published studies may not be included in literature reviews: A – all studies that have been conducted; B – studies that are published; C – studies that are included in literature reviews*

The existence of publication bias was first suspected from the observation that a large proportion of published studies had rejected the null hypothesis. In 1959, Sterling found that the results of 97% of studies published in four major psychology journals were statistically significant, concluding that studies with non-significant results might be under-reported.[8] An early example of the identification of publication bias in medical research is that by Chalmers and Grady,[9] who in 1965 attempted to explain the variability in reported rates of death due to serum hepatitis. It was suspected that there was a tendency for clinicians or editors to publish unusual findings. A search of the MEDLINE database found that, in the medical literature, the term "publication bias" first appeared in a study published in 1979.[10]

Although bias in the published literature may imperil the validity of both traditional narrative reviews and systematic reviews,[11] the problem of the selective publication of studies has been highlighted only recently in medical research, coinciding with an increasing use of meta-analysis and systematic review.[12] The numbers of articles using meta-analysis and about publication bias have increased simultaneously during the 1990s (*Table 1*).

Definition of publication and related biases

Bias is defined as: deviation of results or inferences from the truth, or processes leading to such deviation; or any trend in the collection, analysis,

TABLE 1 *Number of articles about meta-analysis versus number of those with the term "publication bias" or "file drawer" indexed in MEDLINE over different years*

Years	No. found by using term "meta-analy*"	No. found by using terms "publication bias" or "file drawer"
1966–1984	51	2
1985–1989	462	18
1990–1992	1214	31
1993–1995	1587	97
1996–1998	2147	147

interpretation, publication or review of data that can lead to conclusions that are systematically different from the truth.[13] More technically, an estimator t of a parameter T is biased if $E(t) \neq T$, where E denotes the expected value.[14] It should be noted that bias may refer to both the result of a biased process and the biased process itself.

Publication bias has been defined in various ways by different people. For example, Chalmers and colleagues[15] considered publication bias broadly to have three stages: prepublication bias in the performance of research; publication bias; and postpublication bias in interpretations and reviews of published studies. Publication bias was more specifically defined by Dickersin as "the tendency on the parts of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings".[16]

In Dickersin's definition of publication bias, there are two basic concepts: study findings and publication. Study findings are commonly classified as being statistically significant or non-significant. In addition, study results may be classified as being positive or negative, supportive or unsupportive, favourite or unfavourite, striking or unimportant. It should be noted that the classification of study findings is often dependent on subjective judgement and may be unreliable. For example, people may have different understandings about what are negative findings.[17]

The formats of publication include full publication in journals, presentation at scientific conferences, reports, book chapters, discussion papers, dissertations or theses. In fact, "publication is not a dichotomous event: rather it is a continuum".[18] Although a study that appears in a full report in a journal is generally regarded as published, there may be different opinions about whether it should be classified as published or unpublished when results are presented in other formats.

The accessibility of research results is dependent not only on whether a study is published but also on when, where and in what format this occurs. In this review, we have used the term "dissemination profile" to describe the accessibility of research results, or the possibility of research findings being identified by potential users. The spectrum of the dissemination profile ranges from completely inaccessible to easily accessible, according to whether, when, where and how research is published. Dissemination bias occurs when the dissemination profile of a research report depends on the direction or strength of its findings. That is, dissemination bias includes bias due to selective publication and other publication-related biases caused by the time, type and language of publication, multiple publication bias, selective citation of references, database index bias, and biased media attention (see *Box 1* for definitions).

The advantages of the term "dissemination bias" are that it avoids the need to define publication status and it is more directly related to accessibility than publication bias. For example, media attention can have a major impact on dissemination, but it is not normally included within the definition of publication bias. Also, the term "publication bias" is sometimes used to refer not only to bias due to selective publication but also to other biases related to the time, type and language of publication, multiple publications and selective citation of references. Dissemination bias may be a better expression with which to replace this broad use of the term publication bias.

This review concerns dissemination bias in medical and health-related research. Because the non-publication of results poses the most serious difficulty in locating relevant studies, publication bias is its main focus.

The authors present the results of a systematic review of studies that have examined methodological issues or provided empirical evidence concerning publication and related biases. After a description of the methods used (chapter 2), empirical evidence of publication and related biases (chapter 3), consequences (chapter 4) and sources of publication bias (chapter 5) are presented and discussed. The methods for dealing with publication and related biases and their

BOX 1 Definitions of publication and related biases

Dissemination bias: Occurs when the dissemination profile of a study's results depends on the direction or strength of its findings. The dissemination profile is defined as the accessibility of research results or the possibility of research findings being identified by potential users. The spectrum of the dissemination profile ranges from completely inaccessible to easily accessible, according to whether, when, where and how research is published.

Publication bias: Occurs when the publication of research results depends on their nature and direction.[16] The results of published studies will then be systematically different from those of unpublished studies.

Specific types of publication and dissemination bias

Positive results bias: Occurs when authors are more likely to submit, or editors accept, positive than null results.[19]

Hot stuff bias: Occurs when a topic is "hot", and neither investigators nor editors may be able to resist the temptation to publish additional results, no matter how preliminary or shaky.[19]

Time-lag bias (pipeline effect[20]): Occurs when the speed of publication depends on the direction and strength of the trial results.[21] For example, studies with significant results may be published earlier than those with non-significant results.

Grey literature bias: Occurs when the results reported in journal articles are systematically different from those presented in reports, working papers, dissertations or conference abstracts.

Full publication bias: Occurs when the full publication of studies that have been initially presented at conferences or in other informal formats is dependent on the direction and/or strength of their findings.

Place of publication bias: In this review, this is defined as occurring when the place of publication is associated with the direction or strength of its findings. For example, studies with positive results may be more likely to be published in widely circulated journals than studies with negative results. The term was originally used to describe the tendency for a journal to be more enthusiastic towards publishing articles about a given hypothesis than other journals, for reasons of editorial policy or readers' preference.[22]

Outcome reporting bias: Occurs when a study in which multiple outcomes were measured reports only those that are significant.

Multiple publication bias (duplicate publication bias): Occurs when studies with significant or supportive results are more likely to generate multiple publications than studies with non-significant or unsupportive results. Duplicate publication can be classified as "overt" or "covert".[23] Multiple publication bias is particularly difficult to detect if it is covert, when the same data are published in different places or at different times without providing sufficient information about previous or simultaneous publication.

Language bias (Tower of Babel bias[24]): Occurs when languages of publication depend on the direction and strength of the study results.

Citation bias (reference bias, one-sided reference bias[19]): Occurs when the chance of a study being cited by others is associated with its result. For example, authors of published articles may tend to cite studies that support their position. Thus, retrieving literature by scanning reference lists may produce a biased sample of articles and reference bias may also render the conclusions of an article less reliable.[25]

Database bias (indexing bias): Occurs when there is biased indexing of published studies in literature databases.[4] A literature database, such as MEDLINE or EMBASE, may not include and index all published studies on a topic. The literature search will be biased when it is based on a database in which the results of indexed studies are systematically different from those of non-indexed studies.

Retrieval bias: Occurs when there is a difference between the average estimate based on all studies conducted and the average estimate based on studies retrieved in a research domain.[20] The studies retrieved may include published and some unpublished studies.

Media attention bias: Occurs when studies with striking results are more likely to be covered by the media (newspapers, radio and television news).

strengths and limitations are then examined (chapters 6 and 7). The results of a survey of published systematic reviews are also presented to provide further evidence of publication bias and to illustrate the methods used for dealing with it (chapter 8). Finally, we summarise and discuss the major findings of this review and make recommendations for future research (chapter 9).

Chapter 2

Review methods

This report includes a systematic review of publication and related biases, and a survey of publication bias in published systematic reviews.

Systematic review of relevant literature

The aim was systematically to identify and appraise studies that have examined methodological issues and provided empirical evidence about publication bias and related problems. The review focuses on the following questions:

- What empirical evidence is available on the existence and consequences of publication and related biases?
- What are the causes and risk factors of publication and related biases?
- What methods have been developed and how useful are these methods for preventing, detecting and correcting publication and related biases?

Criteria for inclusion and exclusion

A study was considered to be methodological if its main objectives involved any of the following issues: concept, definition, causes, risk factors, existence and consequences of publication bias; and methods for preventing, reducing, detecting and correcting publication bias. Empirical evidence was defined as any observations that could be used to reveal the existence, magnitude and consequences of the publication and related biases. Many studies can be considered as empirical as well as methodological.

It was difficult to define clear and narrow criteria for including studies in this review because of the broad nature of the related issues and the great diversity of relevant studies. Therefore, all studies relevant to publication-related biases were included. However, if the issue of publication-related biases was mentioned only briefly and was not a major topic, they were excluded.

Search strategy

The following databases were searched to identify relevant literature concerning empirical evidence and methodological issues pertaining to publication and related biases: the Cochrane Review Methodology Database, MEDLINE, EMBASE, BIDS, Library and Information Science Abstracts, PsycLIT, Sociofile, ERIC, Dissertation Abstracts, MathSci, British Education Index, SIGLE and ASSIA. The strategies used to search electronic databases are presented in appendix 1. *Table 2* shows the electronic databases searched and the number of potentially relevant records retrieved. The search of electronic databases was initially conducted in June 1997 and updated in September 1998 to identify more recently published literature. The reference lists of the identified articles were checked. We also contacted experts in the field on an informal basis to identify relevant studies.

Study assessment, data extraction and synthesis

The results of searches of electronic databases were checked independently by two reviewers and any disagreements discussed. The full publications of potentially relevant studies were obtained and their relevance was independently assessed by at least two reviewers. Each was classified as a study of methods, a study of empirical evidence, a focused review, an editorial or a letter concerning publication and related biases. The initial search of electronic databases yielded a total of 4913 records, with many duplications, many studies being indexed in several different databases. From these search results, 200 relevant articles were identified, including 64 studies containing empirical evidence, 51 studies concerning methods, 11 reviews, and 74 editorials or letters.

Data from included studies were extracted by one reviewer using a data-extraction form (appendix 2) and checked by another reviewer. The assessment of relevance and data extraction were carried out without masking the authors of studies or the journals in which these studies were published.

Survey of published systematic reviews

By using a sample of systematic reviews, the aim of the survey was to answer the following questions:

TABLE 2 *Databases searched and the number of potentially relevant records retrieved*

Database	Years	No. records
MEDLINE	1993–06/1997	930
	1987–1992	703
	1981–1986	230
	1976–1980	62
	1966–1975	13
BIDS: Science Citation Index	1981–06/1997	296
BIDS: Social Science Citation Index	1981–06/1997	598
EMBASE	1981–06/1997	328
PsycLIT	1967–07/1997	310
Sociofile	1963–06/1997	138
Information Science Abstracts	1966–07/1997	10
ERIC	1966–04/1997	222
Dissertation Abstracts	1986–06/1997	613
MathSci	1940–07/1997	111
British Education Index	1976–03/1997	2
SIGLE (grey literature database)	1980–06/1997	161
ASSIA	1987–06/1997	186

- What methods have been used in published systematic reviews for identifying and correcting publication bias?
- What further empirical evidence on publication bias can be identified from published systematic reviews?

The Database of Abstracts of Reviews of Effectiveness (DARE),[26] produced by the NHS Centre for Reviews and Dissemination, University of York, provides a sample of published systematic reviews that have already been screened for basic methodological quality. Systematic reviews were identified by regular searching of the major health care databases, including Current Contents Clinical Medicine (weekly), MEDLINE, CINAHL, Allied and Alternative Medicine (all monthly), BIOSIS, PsycINFO, ERIC (all yearly), by handsearching key health care journals, and by scanning the grey literature. To be included in the DARE database, a review has to have at least four of the following six criteria: well-defined review question, appropriate literature search, explicit criteria for including individual studies, validity assessment, presentation of study details, and appropriate synthesis of study results. Systematic reviews in the DARE database included individual studies with different designs such as randomised controlled trials (RCTs), non-randomised controlled studies, case-control studies, cohort studies and case reports.

At the end of August 1998, 193 systematic reviews published in 1996 were included in the DARE database. These were selected to examine the issues and methods relevant to publication and related biases. The reviews were independently assessed by two reviewers by using a data-extraction sheet (appendix 3) to collect the following information:

- type of review (qualitative or meta-analysis)
- whether the issue of publication bias was considered
- whether unpublished studies or those published in non-English languages were searched for and included
- any evidence on the existence, extent and consequence of publication bias
- the methods used for dealing with publication bias.

Chapter 3

Research findings and dissemination profiles

This chapter aims to summarise the empirical evidence about the association between research results and their dissemination profiles. Empirical evidence about the impact and consequences of publication and related biases will be summarised in chapter 4.

Evidence about publication bias can be classified as indirect or direct.[27] Indirect evidence includes the observation of a disproportionately high percentage of positive findings in the published literature and larger effect sizes in small studies compared with large studies. This evidence is indirect because factors other than publication bias may also lead to the observed high percentage of positive findings in the published literature and the observed association between sample size and effect size. Direct evidence includes admissions of bias on the part of those involved in the publication process (investigators, editors and reviewers), comparison of the results of published and unpublished studies, and the follow-up of cohorts of registered studies.

In this chapter, both indirect and direct evidence about the existence of publication bias is summarised. Evidence is also presented about other publication-related biases, including those due to time and type of publication, selective outcome reporting, duplicate publication, country and language, database factors, place of publication, selective citation and media attention. Although many different biases that are publication related may be listed and are discussed separately below, it should be noted that the nature of these biases are often the same or similar in terms of their causes and possible impacts.

Indirect evidence

Proportion of significant results in published studies

In 1959, Sterling found that the results of 97% of studies published in four major psychology journals were statistically significant, concluding that those with non-significant results might be under-represented.[8] In 1995, the same author and colleagues concluded that the practices leading to publication bias have not changed over a period of 30 years.[28] Sterling's observation has been confirmed by other authors, who have noted that the proportion of studies with significant or positive results ranges from 35% to 97% of all published studies (*Table 3*).[29–37]

The proportion of studies published in medical journals that report significant or positive results seems to be lower than in psychological journals. A much lower proportion of studies with significant results was observed in the two reports that examined statistical power in published trials that was insufficient to detect small or moderate differences.[35,36] In a survey of 383 RCTs published in three general medical journals, Moher and co-workers[35] found that the percentage of trials having negative results was 33% in 1975, 27% in 1980, 25% in 1985, and 25% in 1990. Another study assessed a sample of 386 RCTs and found that the proportion with a significant result was 17% in 1976, 33% in 1981, 38% in 1986, and 46% in 1991.[36]

This kind of evidence is not reliable for confirming the existence of publication bias because the expected proportion with significant or positive results in all studies (both published and unpublished) is unknown. It has also been argued that the majority of studies may report significant results if the hypotheses to be tested are not selected at random.[38]

Association between sample size and treatment effect

It could be assumed that the true treatment effect of a health care intervention for a defined indication should not be affected by the study sample size if there is no bias. However, the results from smaller studies will be more widely spread around the average owing to greater random error. It is therefore generally believed that small studies are more vulnerable to publication bias.[7] In fact, the assumed association between publication bias and sample size is the cornerstone of many methods for detecting publication bias in meta-analysis. (The association between sample size and publication bias is discussed in more detail in chapter 5.)

By examining a consecutive sample of 246 published cancer clinical trials, Berlin

TABLE 3 *Proportion of published studies with significant results*

Study (year)	% of significant or positive results	Definition and sources of primary studies
Sterling (1959)[8]	97 (286/294)	Research reports that rejected the null hypothesis ($p < 0.05$) in all research reports that used statistical tests; four psychology journals
Sterling *et al.* (1995)[28]	96 (538/563)	Research reports that rejected the null hypothesis ($p < 0.05$) in all research reports that used statistical tests; eight psychology journals
Sterling *et al.* (1995)[28]	85 (270/316)	Research reports that rejected the null hypothesis ($p < 0.05$) in all research reports that used statistical tests; three medical journals
Smart (1964)[29]	91 (282/309)	Positive studies are those in which at least half of the null hypotheses were rejected; four major psychology journals
Smart (1964)[29]	75 (126/169)	Positive studies are those in which at least half of the null hypotheses were rejected; abstracts presented at the American Psychological Association annual meetings, and PhD dissertation abstracts in psychology
Bozarth and Roberts (1972)[30]	94 (841/895)	Research articles that rejected the null hypothesis ($p < 0.05$) in all research reports that used statistical tests; three psychology journals
Greenwald (1975)[32]	88 (175/199)	Studies that rejected the null hypothesis; *Journal of Personality and Social Psychology*
Hubbard and Armstrong (1997)[31]	92 (638/692)	Studies that rejected the null hypothesis ($p < 0.05$) in all studies that used statistical tests; 32 randomly selected issues of three marketing/consumer journals
Davidson (1986)[33]	71 (76/107)	Proportion of trials with results favouring the new treatment; clinical trials from five general medical journals
Moscati *et al.* (1994)[34]	80 142/177	Proportion of studies with positive outcomes (the new treatment is significantly better than the standard treatment); emergency and general medical journals
Moher *et al.* (1994)[35]	73 (281/383)	Trials with positive results (defined according to the explicit statement in the text or the primary outcome measure); from *JAMA*, *The Lancet* and the *NEJM* in 1975, 1980, 1985 and 1990
Mulward and Gotzsche (1996)[36]	35 (136/386)	Significant difference in effect; double-blind RCTs published in 1976, 1981, 1986 or 1991, with active treatments in both arms, were not a crossover design, were published in English as full articles, and which had clinical outcomes; trials were identified through a MEDLINE search
Csada *et al.* (1996)[37]	91 (1098/1201)	The main hypothesis was statistically supported in 1201 papers that used statistical tests; randomly selected from 43 biological journals
JAMA, Journal of the American Medical Association; NEJM, New England Journal of Medicine		

and colleagues[39] found that treatment effects reported in smaller studies were greater than those in larger studies. In addition, they observed that non-randomised trials and single-centre studies tended to be more vulnerable to publication bias than randomised and multi-centre studies. Evidence provided by Berlin and co-workers' study was not conclusive because the trials included were heterogeneous in terms of study participants and interventions. The observed difference in the effect size between the small and large studies may be real if the sample size of trials was determined according to perceived treatment effects.[40]

Such analyses may be more appropriately carried out using homogeneous studies that aim to estimate a treatment effect within a meta-analysis. Allison and colleagues[41] assessed four published meta-analyses of obesity treatment, using a regression technique to assess the association between treatment effect and sample size. A significant association was observed in two meta-analyses that contained

a large number of primary studies (68 and 418 respectively). For the other two meta-analyses, there was no significant association between the treatment effect and sample size, possibly due to insufficient statistical power (13 and 15 primary studies respectively).[41]

The funnel plot is often used to examine visually the association between sample size and treatment effect in a meta-analysis (more discussion about the funnel plot is available in chapter 7). According to an asymmetric funnel plot, Egger and Davey-Smith[42] concluded that publication bias might be blamed for the misleading result from a meta-analysis of intravenous magnesium in myocardial infarction. In another review of 16 trials of antiplatelet agents for intrauterine growth retardation and in proteinuric pre-eclampsia, smaller trials were associated with a greater treatment effect than larger trials.[43] Egger and colleagues[44] found that significant asymmetry existed in 38% of journal-published meta-analyses and in 13% of Cochrane reviews. However, other alternative explanations for the observed association between sample size and treatment effect (such as different intensity of interventions, differences in underlying risk, study design quality and chance) need to be excluded before the existence of publication bias can be confirmed.[44]

Direct evidence

Survey of investigators

Surveys of investigators and authors have provided some evidence on the existence of publication bias and its magnitude (*Table 4*). Questionnaires or letters were sent to authors of published studies,[32,45,46] members of academic or professional organisations,[47–49] or clinicians.[50] According to a survey of authors of articles submitted to a psychological journal, the probability of investigators submitting for publication was 0.49 for studies with statistically significant results but only 0.06 for studies with non-significant results.[32] In another survey of the members of the American Psychological Association it was found that the rate of publication was 66% for 129 positive studies and only 22% for 65 neutral or negative studies.[47]

Dickersin and co-workers[45] observed that the proportion in which the new treatment was better than the control therapy was 14% in 178 unpublished clinical trials and 55% in 767 that were published. A survey of the members of the Society for Menstrual Cycle Research in North America found that the rate of publication was 73% for 30 studies with positive results and 54% for 26 studies with null findings.[48] Rotton and colleagues[46] reported that a non-significant result was the most frequent reason given by authors for not publishing a study.

Although the results of these surveys suggest the existence of publication bias, the response rate ranges from 49% to 79% and it is difficult to assess the reliability of self-report data. The investigators or authors surveyed were often selected non-randomly and thus the results may not be generalisable.

Follow-up of cohorts of registered studies

The existence of publication bias has been consistently confirmed by studies that have retrospectively followed up cohorts of studies approved by the Research Ethics Committee (REC) in the UK or in Australia, the Institutional Review Board (IRB) in the USA or cohorts of trials registered by research sponsors (*Table 5* and *Figure 2*).[51–54] The findings from these studies will be referred to later when relevant; only the main findings are discussed below.

Easterbrook and co-workers[51] retrospectively surveyed 487 studies approved by the Central Oxford REC between 1984 and 1987. Information on the studies was obtained by writing to and telephoning the principal investigators. Studies that had appeared in or been accepted by a journal were considered to be published, but not studies presented in book chapters or meeting abstracts. The study results were classified as "statistically significant" ($p < 0.05$), "non-significant trend" (difference observed but $p \geq 0.05$), or "null" (no difference). For studies that did not use a statistical test, results were classified as "striking", "definite but not striking", or "null". Among 285 of the studies that had been analysed by May 1990, the proportion that had been published was 60% for those with significant results and 34% for those with a non-significant trend or null results.

Dickersin and colleagues[52] retrospectively followed up 737 studies approved before or in 1980 by two IRBs at the Johns Hopkins Health Institutions. The principal investigators of the eligible studies were contacted to obtain information on study characteristics. Significant results included those that were statistically significant or had findings of "great importance" when a

TABLE 4 *Surveys of authors concerning unpublished studies*

Study (date) and survey sample	Findings
Greenwald (1975)[32] 48 authors and 47 reviewers of articles submitted to the *Journal of Personality and Social Psychology* in 1973; 36 authors and 39 reviewers responded	The survey found self-reported evidence of substantial biases against the null hypothesis in formulating a research problem and in deciding what to do with the data once collected. The probability of submitting for publication was 0.49 for studies rejecting the null hypotheses and 0.06 for studies not rejecting the null hypotheses.
Coursol and Wagner (1986)[47] 1000 members of the American Psychological Association; 609 responses	95 unpublished and 99 published outcome studies The rate of submission was 82.2% for 129 positive studies (definition of positive study not provided) and 43.1% for 65 neutral or negative studies. The rate of acceptance was 80.2% for 106 submitted positive studies and 50.0% for 28 submitted negative studies. Overall, the rate of publication was 65.9% for 129 positive studies and 21.5% for 65 negative studies.
Dickersin *et al.* (1987)[45] Questionnaires sent to 318 authors of published trials; 156 responses	271 unpublished; 1041 published Proportion of trials in which the new therapies were better than control therapies: 14% in 178 unpublished trials versus 55% in 767 published trials.
Sommer (1987)[48] 140 members of the Society for Menstrual Cycle Research residing in the North America; 91 responses	28 unpublished; 42 in the pipe-line; 73 published The rate of publication was 73% for 30 studies with positive results, and 54% for 26 studies with null findings. (Positive result was defined as clear menstrual cycle effect in predicted direction.)
Hetherington *et al.* (1989)[50] Initial letters sent to 42 160 obstetricians and paediatricians in 18 countries. Follow-up questionnaires to 481 who indicated having unpublished studies; 453 responses	395 unpublished trials: 18 completed more than 2 years ago; 125 stopped recruitment within 2 years; 193 were actively recruiting and 59 not started The ratio of unpublished to published trials was 1:128, calculated by assuming that 18 unpublished trials were completed between 1940 and 1984, and at least 2300 reports of perinatal trials were published during the same period.
Shadish *et al.* (1989)[49] 519 randomly selected members of five organisations of family and marital psychotherapy; 375 responses	Three unpublished studies; 165 published studies in a meta-analysis The degree of publication bias is still in doubt because of a large number of potential investigators that were not surveyed.
Rotton *et al.* (1995)[46] 740 authors of empirical articles in 75 psychological journals; 468 responses	It was estimated that authors in the sample have "filed away" 15.4% of their work. The most common reason for not publishing was non-significant results (59.9%).

statistical test was not used. The definition of "publication" in this study was broad, including journal articles, monographs, books or chapters in books that were available from medical libraries, or in documents available from a public archive. This survey observed that studies with significant results were more likely to be published than studies with non-significant results (82% versus 66%).[52]

By interviewing the principal investigators, Dickersin and Min[53] obtained information on 217 of the 293 eligible clinical trials funded by the National Institutes of Health (NIH) in the USA in 1979. The methods used to classify the findings and publication status were the same as in the study discussed above.[52] According to the 198 clinical trials completed by 1988, the publication rate was 98% for trials with significant results and 85% for those with non-significant results.

In July 1992, Stern and Simes[54] surveyed the principal investigators of 748 studies approved

TABLE 5 *Publication rate of cohorts of registered studies*

Study	Cohorts of studies	Rate (%) of publication		Adjusted OR (95% CI)[a]	
Easterbrook *et al.* (1991)[51]	Studies approved by the Central Oxford REC between 1984 and 1987	**Published only**		Null	1.00
		Statistically significant	60 (93/154)	Non-significant trend	0.61 (0.23 to 1.59)
		Non-significant trend	35 (12/34)	Statistically significant	2.32 (1.25 to 4.28)
		No difference	34 (33/97)		
		Published or presented			
		Statistically significant	85 (131/154)		
		Non-significant trend	65 (22/34)		
		No difference	56 (54/97)		
Dickersin *et al.* (1992)[52]	Studies approved up to the end of 1980 by the IRBs at the Johns Hopkins Health Institutions	**The School of Medicine and Hospital**		**The School of Medicine and Hospital**	
		Significant results	89 (184/208)	Non-significant	1.00
		Non-significant	69 (93/134)	Significant results	3.55 (1.94 to 6.47)
		The School of Hygiene and Public Health		**The School of Hygiene and Public Health**	
		Significant results	71 (75/106)	Non-significant	1.00
		Non-significant	58 (38/66)	Significant results	1.64 (0.8 to 3.34)
Dickersin and Min (1993)[53]	Completed clinical trials (by 1988) that were funded by the NIH in 1979	Significant	98 (121/124)	Non-significant	1.00
		Non-significant	85 (63/74)	Significant results	7.11 (1.84 to 27.50)
Stern and Simes (1997)[54]	Analysed studies that were submitted to the Royal Prince Alfred Hospital REC between 1979 and 1988	**Quantitative studies**		**Quantitative studies**	
		Significant	68 (99/146)	No difference	1.00
		Non-significant trend	20 (4/20)	Non-significant trend	0.34 (0.17 to 0.67)
		No difference	44 (23/52)	Significant results	2.93 (1.49 to 5.74)
		Qualitative studies		**Qualitative studies**	
		Striking	70 (19/27)	The adjusted ORs were presented in graphs and were not statistically significant	
		Important and definite	59 (35/59)		
		Negative and unimportant	53 (9/17)		

[a] *The adjusted ORs were estimated by using multivariate analyses in which the ORs were adjusted by: study design, study groups, funding source, sample size, importance rating and pilot study or not;*[51] *study design, study group, sample size, funding source and number of centres;*[52,53] *study design and funding source*[54]

between September 1979 and December 1988 by the Royal Prince Alfred Hospital Ethics Committee in Australia. The results of the quantitative studies were classified as "significant" ($p < 0.05$), "non-significant trend" ($0.05 \leq p < 0.10$), or "null" ($p \geq 0.10$). The results of qualitative studies were classified subjectively by the principal investigators as "striking", "important and definite", or "unimportant and negative". Studies were considered as being published if they had appeared in a peer-reviewed journal. Questionnaires were completed for 520 of the 748 eligible studies. It was found that the publication rate was 68% for studies with significant results, 20% for those with a result of non-significant trend, and 44% if there was a null result. For the 103 qualitative studies, the publication rate was 70% for those with striking results, 59% for those with important and definite results, and 53% when there were unimportant and negative results.[54]

Bias in favour of publication when there were significant results was confirmed in these studies by multivariate analysis adjusting for other study characteristics (*Table 5*). Dickersin combined the adjusted odds ratios (ORs) from these cohort studies and found that the overall adjusted OR for publication bias was 2.54 (95% confidence interval (CI) 1.44 to 4.47).[55] In two of the four cohort studies it was observed that quantitative studies with a non-significant trend were less likely to be published than those demonstrating no difference.[51,54]

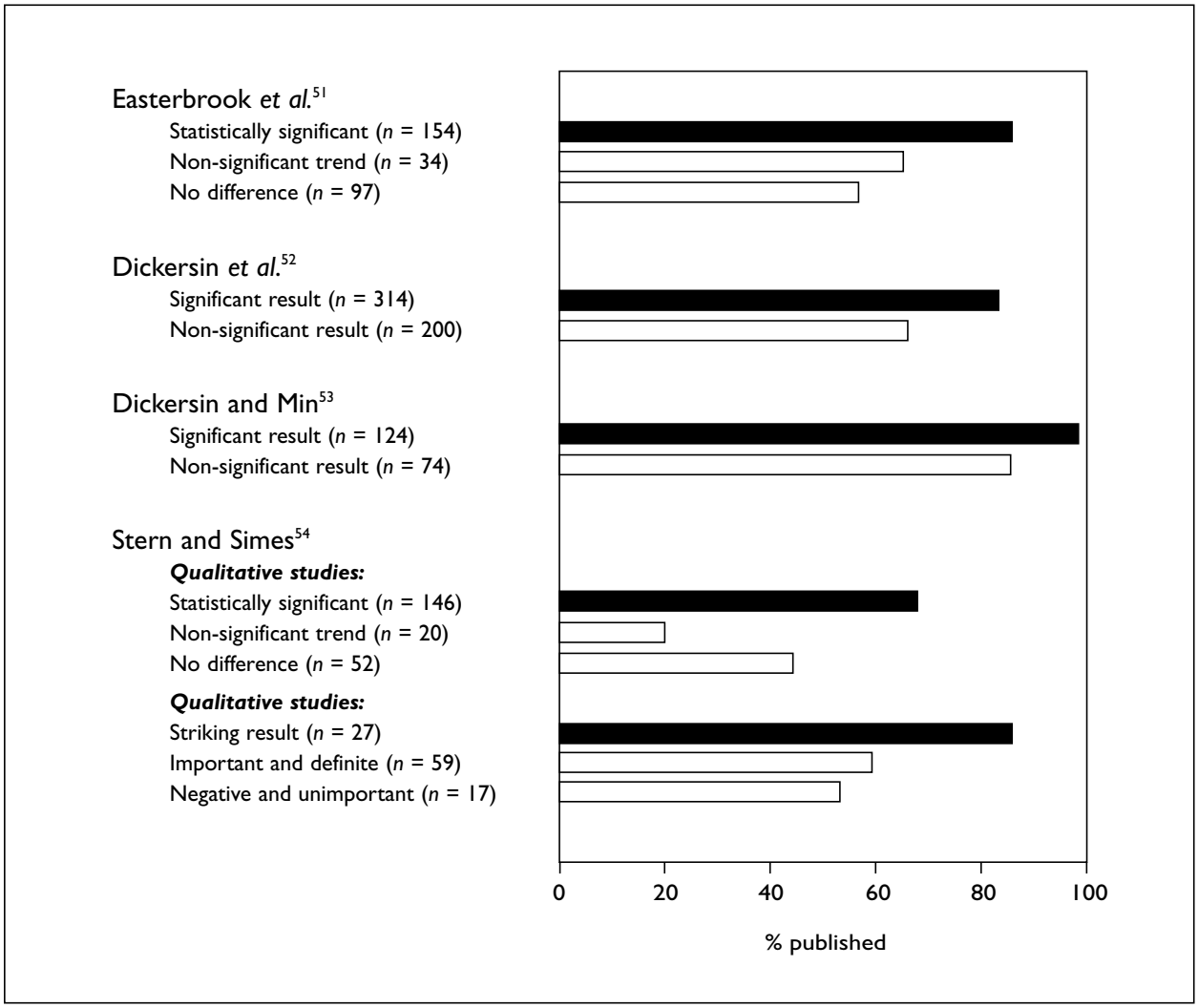


FIGURE 2 *Percentage of studies published and significance of results in registered studies*

Comparison of unpublished with published results

The most direct and convincing evidence of publication bias is from comparisons between unpublished and published results of studies that evaluate the same interventions. Simes compared the results of meta-analyses based on published or registered trials (both published and unpublished) that evaluated combination chemotherapy versus an initial alkylating agent in advanced ovarian cancer.[11,56] When only published trials were pooled, the overall survival rate was significantly higher in the combination chemotherapy group than in the initial alkylating agent group (median survival ratio 1.17; $p = 0.01$). However, the survival advantage for combination chemotherapy was not observed when all registered trials were pooled (median survival ratio 1.07; $p = 0.09$).

In this example, trials were classified into three publication categories: (A) nine trials that were published but not registered by October 1983; (B) eight trials registered and published by October 1983; and (C) four trials registered but not located by searching MEDLINE and references of retrieved articles. The meta-analysis of published trials included trials belonging to categories A and B, while the meta-analysis of registered trials included trials belonging to B and C. It should be noted that the trials in category C included one that was published in 1984, and one that was published in 1980 but not identified by the initial literature search.[11,56] If all the trials published (including those published after October 1983 and those not identified by the initial search) were combined, the pooled OR was 1.13, which was still greater than that of registered studies (OR = 1.07).

Using the same method, Simes also compared registered trials with published trials that evaluated the survival advantage of combination

chemotherapy versus a single alkylating agent plus prednisone in multiple myeloma.[56] Meta-analysis of registered trials yielded a smaller survival benefit of combination chemotherapy when compared with meta-analysis of published trials (median survival ratio 1.22 versus 1.66).

Studies that are presented in reports, dissertations and conference proceedings may sometimes be considered as unpublished. More evidence from the comparison of published studies and unpublished studies (including those published informally) will be discussed when grey literature bias is discussed.

Other publication-related biases

The accessibility of research results is dependent not only on whether a study is published but also on when and in what format. Other biases related to the dissemination of research results include those concerning time lag, grey literature, outcome reporting, duplicate publication, language, database, citation and media attention. The available evidence about these and other biases is discussed below.

Time lag bias

When the speed of publication depends on the direction and strength of the trial results, this is referred to as time lag bias.[21] In a survey of studies approved by a hospital ethics committee in Australia,[54] time to publication was defined as the time from approval by the ethics committee to first publication in a peer-reviewed journal. Using this definition, studies with significant results were published much earlier than those with null results (median 4.8 years versus 8.0 years; hazard ratio 2.32; 95% CI 1.47 to 3.66). Adjusting for other predictors of publication (research design and funding source) did not change this result materially. When only the large quantitative studies (sample size > 100) were analysed, time lag bias was still evident (hazard ratio 2.00; 95% CI 1.09 to 3.66).

Further evidence for time lag bias was from a cohort of 109 RCTs, conducted during the period 1986 to 1996, funded by the Division of AIDS of the National Institutes of Allergy and Infectious Diseases in the USA.[57] In this study, results were classified as "positive" if an experimental therapy for AIDS was significantly ($p < 0.05$) better than the control therapy. "Negative results" included those with no statistically significant difference and those in favour of the control therapy. Publication was defined as trial findings being published in a peer-reviewed journal. The median time from starting enrolment to peer-reviewed publication was 6.5 years for negative trials and 4.3 years for positive trials. The hazard ratio for time to publication for positive versus negative trials was 3.7 (95% CI 1.8 to 7.7). The median time to first submission after completion was 1.0 year for positive trials versus 1.6 years for negative trials ($p = 0.001$). The median time to publication after submission was 0.8 versus 1.1 years respectively for positive and negative trials ($p = 0.04$).

Misakian and Bero[58] contacted 89 organisations that provided funding for passive smoking research. Time to publication was measured by using the start date of funding because it was difficult to decide the time of completion. Published studies were those that appeared or were in press in a peer-reviewed or non-peer-reviewed publication, but not those published only as conference abstracts. From 65 respondents to a semistructured telephone interview, they identified 61 studies funded between 1981 and 1995. The median time to publication was 5 years for statistically non-significant studies and 3 years for statistically significant studies. Multivariate analysis revealed that the time to publication was associated with statistically significant results ($p = 0.004$), experimental study design ($p = 0.01$), study size less than or equal to 500 ($p = 0.01$) and animals as subjects ($p = 0.03$).

Rothwell and Robertson[59] found that the treatment effect was overestimated by early trials compared with the subsequent trials in 20 of 26 meta-analyses of clinical trials. The average difference in relative odds was 35% (95% CI 15 to 55).[59] In another study, a significant association ($p < 0.10$) between the year of publication and the treatment effect was observed in four of the 30 meta-analyses that were published in the *British Medical Journal* (*BMJ*) or *JAMA* during the period 1992–1996 and which presented summary data for individual studies.[12]

Grey literature bias

Reports, working papers, dissertations and conference abstracts often have very limited dissemination and are therefore often termed "grey literature".[60] The distinction between the grey literature and unpublished or published studies may be ambiguous. Studies presented in the form of grey literature may be considered as published or as unpublished, according to different definitions.

In 1964, Smart randomly selected 37 theses from *Dissertation Abstracts in Psychology* and found that those with positive results were more likely to be published formally than those with negative results.[29] In the fields of psychological and educational research, several authors observed a tendency for the average effects reported in journal articles to be greater than the corresponding effects reported in dissertations.[61–63]

In a meta-analysis of perioperative parenteral nutrition for reducing complications from major surgery and fatalities, Detsky and co-workers found that the results of studies presented as abstracts reported greater effectiveness than those published as papers.[64] The pooled mean difference in fatality rate was 0.046 ($p = 0.21$) based on the results of published papers but this became greater and statistically significant (0.079; $p = 0.03$) if the results presented in three abstracts were included. In this example, on average, published studies reported smaller treatment effects than conference abstracts.

Devine presented an abstract that compared published studies (from journals or books) with unpublished studies (theses or dissertations) in two meta-analyses.[65] In a meta-analysis of 80 published studies and 102 dissertations concerning surgical patients, published studies yielded larger average estimates of effect than dissertations. Thirty-five published studies and 43 dissertations relating to cancer patients were included in another meta-analysis in which published studies yielded larger average estimates of effect on pain, anxiety and nausea but smaller estimates of effect on vomiting, depression and knowledge when compared with dissertations.[65]

A study in the USA compared data from published studies with data from Food and Drug Administration (FDA) New Drug Application Reviews for assessing non-steroidal anti-inflammatory drug (NSAID)-associated dyspepsia.[66] The quality of unpublished data from FDA reviews was comparable with that of published data. However, the FDA review data suggested that the use of NSAIDs was not associated with dyspepsia (relative risk (RR) 1.1; 95% CI 0.7 to 1.6), while published data showed a significant association between the use of NSAIDs and dyspepsia (RR 1.5; 95% CI 1.2 to 1.8).[66]

McAuley and colleagues evaluated the impact of grey literature on meta-analyses. In this study grey literature was defined as that which is "difficult to identify and retrieve"; it included unpublished studies as well as conference abstracts, theses and industrial reports.[67] From a sample of 135 meta-analyses, they identified 38 that included grey literature. The estimated effectiveness of interventions was on average increased by 12% ($p < 0.05$) when grey literature was excluded.[67]

Full publication bias

A conference abstract can present only very limited data and its accessibility is often restricted. The full publication of research initially presented as abstracts in meetings or journals has been assessed in many studies across a wide range of clinical specialties (*Table 6*).[68–86] The rate of full publication ranged from 23% to 81%. In 1994, Scherer and co-workers[68] combined results from 11 studies and found that, on average, about half the abstracts were published in full after more than 12 months.

Several studies have assessed the association between study outcome and full publication, using different methods to classify outcomes as positive versus negative or neutral, or significant versus non-significant.[68–75] DeBellefeuille and colleagues[69] found that submitted oncology abstracts that had positive results were more likely to be presented (60% versus 35%; $p = 0.03$) and published in full (74% versus 32%; $p < 0.01$) than those with negative results. However, no statistically significant association between study outcome and full publication was observed in seven other studies.[68,70–75]

Koren and co-workers[87] identified 58 abstracts on cocaine use and pregnancy outcome, submitted to the annual meetings of the Society for Paediatric Research between 1980 and 1989. Positive abstracts were defined as those that reported adverse pregnancy outcomes and negative abstracts as those that reported no effect. It was found that submitted abstracts with positive results were more likely to be accepted for presentation than submitted abstracts with negative results (57.1% versus 11.1%; $p = 0.013$), although the design quality of the negative abstracts was similar or better than that of the positive abstracts.

In a study of abstracts submitted to an emergency medicine meeting, Callaham and colleagues[88] reported the results of logistic regression, which showed that the best predictors of acceptance for presentation were a subjective "originality" factor (OR = 2.07; 95% CI 1.13 to 3.89) and positive results (OR = 1.99; 95% CI 1.07 to 3.84). For a submitted abstract to be published, the best predictors were meeting acceptance

TABLE 6 *Rate of full publication of abstracts presented at academic meetings or published in journals*

Study	Specialty: type of abstracts	Method for identifying full publications	Follow-up (months)	% Full publication	Study findings and % full publication
Scherer *et al.* (1994)[68]	Ophthalmology: abstracts of RCTs	Contacting authors and literature search	36	66 (*n* = 93)	Significant 72 (*n* = 46) Non-significant 60 (*n* = 47)
DeBellefeuille *et al.* (1992)[69]	Oncology: a sample of submitted abstracts	Contacting authors and literature search	66	58 (*n* = 197)	Positive 74 (*n* = 65) Negative 32 (*n* = 31) Neutral 56 (*n* = 101)
Chalmers *et al.* (1990)[70]	Perinatology: summary reports of controlled trials	Searching the Oxford Database of Perinatal Trials	48	36 (*n* = 176)	Positive 33 (*n* = 98) Neutral/negative 41 (*n* = 78)
Loep and Kleijnen (1999)[71]	Abstracts of clinical trials, initially published in the *Netherlands Journal of Medicine*	Contacting authors and literature search	> 12	81 (*n* = 131)	Positive 81 (*n* = 89) Negative 81 (*n* = 42)
Landry (1996)[72]	Burn research: presented abstracts	Literature search	60	26 (*n* = 168)	Positive 96 (*n* = 25) Negative 84 (*n* = 107)
Cheng *et al.* (1998)[73]	Cystic fibrosis: conference abstracts	Literature search	< 60	32 (*n* = 178)	No significant association between study outcome and time to publication ($p = 0.54$)
Weber *et al.* (1998)[74]	Emergency medicine: submitted abstracts	Literature search	60	46 (*n* = 492)	Study characteristics (including positive results) did not predict attempts to publish research
Petticrew *et al.* (1999)[75]	Social medicine: presented abstracts	Contacting authors and literature search	< 24	58 (*n* = 77)	No significant association between study outcome and full publication (RR = 0.97; 95% CI 0.60 to 1.57)
Goldman and Loscalzo (1980)[76]	Cardiology: a sample of submitted abstracts	Literature search	37	50 (*n* = 276)	Not available
Corry (1990)[77]	Dental research: a sample of presented abstracts	Literature search	48	23 (*n* = 275)	Not available
Juzych *et al.* (1991)[78]	Ophthalmology: a sample of presented abstracts	Literature search	50	60 (*n* = 175)	Not available
McCormick and Holmes (1985)[79]	Paediatrics: presented abstracts	Literature search	36	48 (*n* = 355)	Not available
Meranze *et al.* (1982)[80]	Anaesthesia: presented abstracts	Literature search	27	33 (*n* = 441)	Not available
Maxwell (1981)[81]	Oncology: presented abstracts	Contacting authors and literature search	12	32 (*n* = 171)	Not available
Dudley (1978)[82]	Surgery: presented abstracts	Literature search	36	57 (*n* = 51)	Not available
Agustsdottir *et al.* (1995)[83]	Oncology: submitted abstracts	Contacting authors	> 12	38 (*n* = 237)	Not available
Yentis *et al.* (1993)[84]	Anaesthesiology: a sample of meeting abstracts published in four journals	Literature search	60	50 (*n* = 215)	Not available
Elder and Blake (1994)[85]	Family medicine and primary care: presented abstracts	Literature search	48	48 (*n* = 475)	Not available
Gavazza *et al.* (1996)[86]	Surgery of the hand: presented abstracts	Literature search	> 36	44 (*n* = 376)	Not available

(OR = 2.49; 95% CI 1.49 to 4.35) and large sample size (OR = 2.26; 95% CI 1.23 to 4.31). They found that the mean effect size was 0.71 for all submitted abstracts, 0.92 for accepted abstracts, and 0.96 for those published. These workers concluded that positive-outcome bias was evident according to the funnel plot of 122 abstracts of prospective studies. However, it is not clear how the signs (direction) of the effect sizes were decided when different interventions were compared. A negative effect size will become positive if the intervention and control groups of a study were defined differently.

Chokkalingam and co-workers[89] compared 53 abstracts with their corresponding full reports on vision and ophthalmology. They found some disagreements in the data presented between the abstracts and the corresponding full publications, although the differences in values were not large. The full publication reported a different number of participants randomised in 13 of the 39 abstracts that presented this data. Three of the 12 results on dichotomous outcomes presented in abstracts disagreed with those in the full reports. Seven of the 15 studies comparing changes in intra-ocular pressure reported a different mean change in the abstract to that in the full article.[89] Another study also observed differences between information presented at final publication and abstracts of RCTs.[90]

Place of publication bias

In a letter to the Editor, Ben-Shlomo and Davey-Smith reported that the *BMJ* published more articles supporting the "early life hypothesis" (about the impact of early life development on the risk of adult disease) than *The Lancet*.[22] They suggested that there may be a "place of publication" bias because, for reasons of editorial policy or readers' preference, one journal is more enthusiastic towards publishing articles about a given hypothesis than other journals.

In a study that compared published and registered trials in advanced ovarian cancer, Simes found that positive trials (indicating a significant survival difference) appeared in prominent journals such as the *NEJM* and *Cancer*, while less widely circulated journals published only negative trials.[11]

Bero and colleagues[91] compared 297 symposium articles in journal supplements and a random sample of 100 journal articles on environmental tobacco smoking published between January 1965 and March 1993. The proportion of review articles was 41% for the symposium articles and 10% for journal articles. It was found that:

> "symposium articles were more likely to agree with the tobacco industry's position (46% vs. 20%), less likely to assess the health effects of ETS [environmental tobacco smoking] (22% vs. 49%), less likely to disclose their source of funding (22% vs. 60%), and more likely to be written by tobacco industry-affiliated authors (35% vs. 6%) than journal articles".

Outcome reporting bias

Outcome reporting bias happens when studies with multiple outcomes report only those that are significant. In many meta-analyses, studies were included only if sufficient data about the relevant outcome were available. If outcome reporting bias exists, the results of meta-analyses may be biased by including an unrepresentative sample of studies.

In a meta-analysis of psychological rehabilitation after myocardial infarction,[92] mortality was significantly lower in the rehabilitation group than in the usual care group (RR 0.65; 95% CI 0.46, to 0.91), according to the results of eight trials that reported total mortality. By contacting the principal investigators of three other trials that did not report total mortality, data were received and the revised RR became 0.73 with a less significant CI (95% CI 0.53 to 1.00). No difference in mortality was observed in a subsequent large multicentre trial (RR 1.01; 95% CI 0.75 to 1.37).[92]

It may be assumed that multiple outcomes measured in a study are associated with the risk of selective reporting of significant results. In a survey of 45 clinical trials published in three general medical journals, Pocock and co-workers[93] found that the median number of end-points was six per trial. They also discussed the risk of the selective reporting of results and other statistical problems such as subgroup analyses, repeated measurements over time, multiple treatment groups, and the overall number of significant tests. It was suggested that "the reporting of clinical trials appears to be biased toward an exaggeration of treatment differences".[93]

Thirty-two RCTs relating to oncology, published in 1992 in the *NEJM* and the *Journal of Clinical Oncology*, were used to assess potential false-positive results due to multiple tests of statistical significance.[94] The median number of therapeutic end-points per trial was five (range 2–19) and 13 trials did not define their primary

end-point. Each of these 32 trials, on average, reported six (range 1–31) statistical comparisons of major outcome parameters. In addition, more than half of the implied statistical comparisons had not been reported. It was concluded that multiple significance testing, publication bias and the low expectation of therapeutic advances all contribute to the probability of reporting false-positive results.[94]

Duplicate (multiple) publication

It has been estimated that 10–25% of the published literature in biomedical sciences represent redundant publications.[95] Multiple publication of the same data in different journals has been condemned mainly for wasting journal space and editors', referees' and readers' time.[96–100] On the other hand, it is arguable that publication of the same data in different ways may be helpful in the dissemination of important research results, providing any previous or parallel publications have been explicitly referenced. The unacceptable "repetitive" publication needs to be distinguished from the necessary "parallel" publication.[101] Therefore, duplicate publication can be classified as "overt" or "covert".[23] Overt duplicate publication is defined as re-analysis of an important trial with appropriate cross-referencing of original reports. Covert duplicate publication is when the same data are published in different places or at different times without adequate reference to a previous or parallel publication.

By examining 44 multiple publications of 31 controlled trials of NSAIDs in rheumatoid arthritis, Gotzsche found important differences in design, exclusion of protocol violators, number of effect variables, number of side-effects, and the significance levels between duplicated publications of the same studies.[102] The conclusion became more positive for the new drugs in the late publications of three trials. He also suggested that multiple publication was difficult to detect because the first author and the number of authors cited are often different.[102]

In a survey of studies approved by an REC, Easterbrook and colleagues[51] found that studies with significant results were more likely to generate multiple publications and more likely to be published in journals with a high citation impact factor when compared with those with non-significant results. A review of RCTs of infertility treatment found that "six studies with a significant result (but none with a nonsignificant result) were reported in four publications from the same institution".[103]

Huston and Moher[104] found that identifying the disaggregation in the multicentre trials of risperidone for schizophrenia was far from obvious because of the chronology of publications, changing authorship, lack of transparency in reporting, and the frequent citation of abstracts and unpublished reports. For example, a North American trial had been reported in part, transparently, and not so transparently, in six different publications by using different author names. It had also been cited in several unpublished forms.

Tramer and co-workers[23] assessed the impact of duplicate data on efficacy estimates of ondansetron on postoperative emesis in a meta-analysis. It was found that, for three trials that were published in six reports, there was no cross-referencing. The estimated number-needed-to-treat (NNT) to prevent one vomit within 24 hours was 9.5 (95% CI 6.9 to 15) in the 16 non-duplicated reports and 3.9 (95% CI 3.3 to 4.8) in the three reports that were duplicated. The efficacy was overestimated by including duplicated data (NNT = 4.9; 95% CI 4.4 to 5.6) compared with the report without duplicated data (NNT = 6.4; 95% CI 5.3 to 7.9). Tramer and colleagues[23] also discussed difficulties in identifying duplicated publications of the same trial data. For example, the same trial might report a different number of patients or different patient characteristics, or use completely different authors in separate publications.

Language bias

Most of the prestigious international journals are published in English, but writing for these journals may be a problem for researchers who are non-native English speakers.[105] A story that appeared in *Nature* reported that a French scientist had no chance of being promoted because his work had not been published in English.[106]

By assessing several issues of *Current Contents in Life Science*, Henrissat found that the number of publications per million inhabitants was 62 in California, 41 in the UK, 39 in Canada, 31 in Australia, 38 in Sweden, 17 in France, 12 in West Germany, 11 in Japan, and ten in Italy.[107] It was concluded that English-speaking countries generally produced more publications per capita than non-English-speaking countries, which might be caused by difficulty in writing in English for other language speakers and the domination of English-speaking scientists on the editorial boards of international journals.

However, Braun and Schubert[108] argued that "the existence of national publication bias is a far more complex and multidimensional problem than inferred from somewhat simplistic approaches" used by Henrissat.[107] They examined the number of publications originating from selected countries using "1981–1985 Scientometric Datafiles" and found that the numbers of publications per million inhabitants in Sweden, Israel, Denmark and Switzerland were greater than those in the UK, Canada, USA or Australia.[108]

Gregoire and colleagues[24] examined meta-analyses in eight medical journals published in English between January 1991 and April 1993. They found that 28 of the 36 meta-analyses had language restrictions. By using the computerised search strategies reported in these 28 meta-analyses, they identified 19 individual studies that had not been included for linguistic reasons. The inclusion of eight of these 19 studies made no difference in five corresponding meta-analyses, while the inclusion of the other 11 studies had the potential to modify the seven corresponding meta-analyses. The most important difference was the change in the 95% CI of the overall OR estimated in a meta-analysis of selective decontamination of the digestive tract in intensive care units. The pooled OR was 0.70 (95% CI 0.45 to 1.09) in the original meta-analysis and became 0.67 (95% CI 0.47 to 0.95) after including a study published in a Swiss journal.[24] This result suggested that treatment effects may be underestimated if studies published in non-English languages are excluded.

It was found that Scandinavian referees awarded higher quality scores to a manuscript written in English than to a manuscript in a referee's own national language, although the methodological flaws were identical in the two fictitious manuscripts used.[109] Egger and co-workers[110] identified 40 pairs of RCTs, each pair consisting of an RCT published in German and a matched RCT by the same author published in English during the same period. They found that design characteristics and quality features were similar between RCTs published in German and RCTs conducted in German-speaking Europe but published in English. Statistically significant results ($p < 0.05$) were reported in 35% of German language articles and 62% of English language articles (OR = 3.75; 95% CI 1.25 to 11.3). Logistic regression analysis found that a significant result was the only variable that was associated with a trial's publication in English language journals. It was concluded that "authors were more likely to publish RCTs in an English language journal if the results were statistically significant".[110]

Country bias

There is some evidence concerning variable results between different countries. Ottenbacher and Difabio observed that the estimated efficacy of spinal manipulation therapy was greater in studies reported in English language journals published outside the USA than for similar studies in journals published in the USA (average effect size 0.45 versus 0.29).[111] It was suggested that this finding might be explained by the existence of publication bias and/or other intervention characteristics.

In a study of abstracts of trials from MEDLINE, Vickers and colleagues[112] found that the proportion of positive results in trials comparing acupuncture with controls was 100% for trials originating from China, Taiwan, Japan and Hong Kong; on average, it was 56.7% from 14 western developed countries such as the USA, Sweden, UK, Denmark, Germany and Canada. They also found that the percentage of positive or favourable results in trials of interventions other than acupuncture was 99% for trials originating from China, 97% from the USSR/Russia, 95% from Taiwan, 89% from Japan, and 75% from England. It was concluded that publication bias is a possible explanation for the unusually high proportions of positive results reported from some countries.[112] The existence of publication bias in Chinese journals of traditional medicine has been indicated by an asymmetric funnel plot of 49 trials of acupuncture in the treatment of stroke.[113]

Database bias

There may be a tendency to publish studies with negative or less positive results in low-circulation journals that are not indexed in commonly used electronic databases.[114] It was estimated that about 98% of journals indexed in the major literature databases (such as MEDLINE or the Science Citation Index) were from western developed countries and only 2% from less developed countries.[115] Nieminen and Isohanni suggested that there is a bias against European journals in medical literature databases because 27% of 320 psychiatric research articles by Finnish authors and published in English were not included in the MEDLINE database.[116]

It has been confirmed consistently that there is a greater possibility of missing relevant studies by searching a single electronic database.[117–124] In a study of 814 references of included studies in 75 Cochrane neurological reviews, Taus and colleagues found that 79% of these references were from journals indexed in MEDLINE or

EMBASE and 21% were from other journals or grey literature.[125]

The retrieval of relevant studies using electronic databases is dependent on journal coverage of these databases. A study found that more than 400 of 977 psychiatry journals were not indexed in any of the four commonly used databases (PsycLIT, EMBASE, BIOSIS and MEDLINE) and there was a high proportion (35%) of journals indexed in only one of these four databases.[126] However, there is a lack of empirical evidence about database bias and indexing bias in terms of study results on a given topic.

Citation bias

The motivation to cite references has been classified as: persuasiveness (to convince the correctness of the methods and results), positive credit (to give positive credit to the material referenced), currency scale (to show how up to date they are), reader alert (to alert the reader to new, different or obscure references), operational information (to borrow methods and techniques), social consensus (to be considered important by colleagues), and negative credit (to criticise, correct, disclaim and dispute other studies).[127] Shadish and colleagues randomly selected one citation from each of 283 articles published in three psychological journals and asked each author about the most important reason for citing the selected reference.[128] It was found that citation was most commonly used to support the author's argument. Study quality was not an important consideration in 98% of cases.

In one study examining the judgement and decision literature, it was found that negative results were significantly more likely to be cited than positive results. This could not be explained by the journal's popularity or the year of publication.[129] However, it was claimed that the citation bias observed in this study was questionable because the two types of articles were published in different types of journals and reported different kinds of evidence.[130]

Possible citation bias was examined by using 111 comparative trials on NSAIDS.[25] The pattern of citations in these trials was classified as positive, neutral or negative selection of references by comparing the proportion of positive and negative references. For example, it was classified as positive when the proportion of trials with a positive outcome in the reference list was higher than that in all available trials. Among 76 trials in which citation bias was possible, the selection of references was classified as neutral in ten, negative in 22, and positive in 44. Therefore, positive selection of references is more likely to happen than neutral and negative selection.[25]

By comparing the citation frequency of cholesterol-lowering trials, it was found that supportive trials were cited almost six times more often than others (the mean annual number of citations was 40 versus 7.4).[131] This difference in citation frequency could not be explained by the type of journal and the sample sizes of trials. Another study examined quotations in three authoritative reviews on diet–heart issues and found that only one of six relevant RCTs with a negative outcome was cited and by only one of the three reviews.[132] On the other hand, two, four and six non-randomised trials with a positive outcome were cited in each review respectively. It was suggested that "fundamental parts of the diet–heart idea are based on biased quotation".[132]

Hutchison and co-workers[133] assessed citation bias by comparing the proportion of relevant supportive and non-supportive trials used in 17 reviews on the clinical effectiveness of pneumococcal vaccine. Supportive trials were defined as those that reported significantly fewer failures in vaccinated subjects than among the controls. It was found that unsupportive trials were more likely to be cited than supportive trials (11.9% versus 5.8%). The tendency to cite recent trials may be one reason for this disproportionate citation of unsupportive studies because six of the seven trials published after 1980 were unsupportive and all seven trials published before 1980 were supportive.[133]

In an assessment of published reviews on the prophylactic removal of impacted third molars, it was found that reviews with similar aims included very different evidence on which to draw conclusions.[134] Of 69 studies that were used in nine general reviews about the association between pathology and impacted third molars, one was quoted in five reviews while 43 were cited only once. This discrepancy in the use of relevant studies cannot be reasonably explained by the year of publication or quality criteria.

Joyce and colleagues[135] found that only three of 89 reviews of chronic fatigue syndrome described the search method, and that the discipline and nationality of the review authors influenced citation of the literature. Seventy-two per cent of references cited by authors from

the USA were published in the US journals, while 55% of references cited by authors from the UK were in the UK journals ($p = 0.001$). Authors with laboratory-based disciplines were more likely to cite laboratory references, while authors from the psychiatric field were more likely to cite the psychiatric literature ($p = 0.01$).

By searching MEDLINE from 1966 to August 1997, Budd and co-workers[136] identified 235 studies that had been retracted because of acknowledged errors, results that could not be replicated, misconduct, or unknown reasons. There was a total of 2034 citations to the 235 retracted articles after the retraction notice, according to a search of Science Citation Index. After analysing 299 of those citations that were in journals in the *Abridged Index Medicus*, it was concluded that "retracted articles continue to be cited as valid work in the biomedical literature after publication of the retraction".[136]

Media attention bias

Media attention bias exists because:

> "the science or medical journalist has to search for news offering headlines that are interesting not only to the potential reader, which is the over-riding objective of any journalist, but also to their own section colleagues and even more so to the person in charge of the decision to publish the story and position it in the newspaper."[137]

Consequently,

> "news of killer bacteria, exterminating viruses, and miraculous therapies tends to have greater appeal because such stories compete with murders, rapes, ecological catastrophes, and declarations from famous people."[137]

Combs and Slovic[138] found that the coverage by newspapers about causes of death was not related to the statistical frequency of their occurrence. The newspapers overemphasised homicides, accidents and disasters, and underreported diseases as causes of death. Violent accidents and homicides make more interesting and exciting stories than diseases. It was suggested that:

> "the biased selection of newspaper reports may serve some useful function by alerting society to the need for correcting action against hazards that, if neglected, might cause the premature death of many persons."

It was not clear whether media coverage was influenced by people's opinions about what is important, or whether people's judgements were influenced by the media coverage, although both of the two directions of influence are possible.[138]

Popular press coverage of research on the association between alcohol and breast cancer was examined by Houn and colleagues.[139] They identified 58 scientific articles and 89 newspaper or magazine stories. Only 11 of these 58 scientific articles were cited in the newspaper or magazine stories. Press stories cited all scientific articles that were published in *JAMA* and the *NEJM*. Articles published in journals other than *JAMA* and the *NEJM* were often ignored by the newspaper and magazine reports. There was no significant difference between the scientific articles and press stories in the frequency of reporting positive, negative or neutral results. It was concluded that "the vast majority of scientific studies on alcohol and breast cancer were ignored in press reports".[139]

Koren and Klein[140] compared newspaper coverage of one positive study[141] and one negative study[142] on radiation as a risk for cancer, published in the same issue of *JAMA* in 1991. Nine of the 19 newspaper reports covered only the positive study. In the other ten reports that covered both the positive and the negative study, the average number of words was 354 for the positive result and 192 for the negative result. It was suggested that the number, length and quality of newspaper reports on the positive study were greater than news reports on the negative study, which suggests a bias against news reports of studies that show no effects or no adverse effects.[140]

Summary

The evidence concerning the association between dissemination profile and study findings are presented. It should be stressed that there may be problems in the definition of publication status and the classification of study findings between the studies reviewed. In addition, studies of publication bias themselves may be as vulnerable as other studies to the selective publication of significant or striking findings.[7] The available evidence therefore needs to be interpreted with caution.

The existence of publication bias has been demonstrated by direct evidence about the association between publication and study finding. In terms of other publication-related biases, there is direct evidence indicating that studies with significant results are often

published earlier than those with non-significant results. However, limited and often indirect evidence indicates only the possibility of full publication bias, outcome reporting bias, duplicate publication bias, database bias and language bias. There is some direct evidence about the existence of citation bias and media attention bias. Although the extent and direction of publication and related biases are uncertain and may vary greatly depending on the circumstances, it seems reasonable to conclude that studies with significant or favourable results are more widely disseminated than those with non-significant or unfavourable results.

Chapter 4

Consequences of publication and related biases

Evidence summarised in chapter 3 suggests that the dissemination profile of research projects may be associated with the strength or direction of their results. As a direct consequence of publication and related biases, the estimates of treatment effects or associations between variables may be misleading if only published studies are included in systematic reviews and meta-analyses. The treatment effect of a health care intervention may be overestimated if studies with positive results are more likely to be published. For example, a meta-analysis of registered studies failed to show the significant survival benefit observed in a meta-analysis of published studies of combination chemotherapy versus initial alkylating agent in advanced ovarian cancer.[11] It is also possible that the treatment effect may be underestimated because of publication and related biases. For example, a meta-analysis underestimated the treatment effect of selective decontamination of the digestive tract in intensive care units because a study was not included for the reason of it being published in a non-English language journal.[24]

The possible consequences of publication and related biases are illustrated in *Figure 3*. The empirical evidence on misleading information gained from biased literature reviews has been summarised in chapter 3. Chapter 4 now focuses on the impact of publication and related biases on health policy, clinical decisions and outcomes of patient management. Publication and related biases may result in avoidable suffering of patients and the waste of limited resources. In addition, as a consequence of these biases,

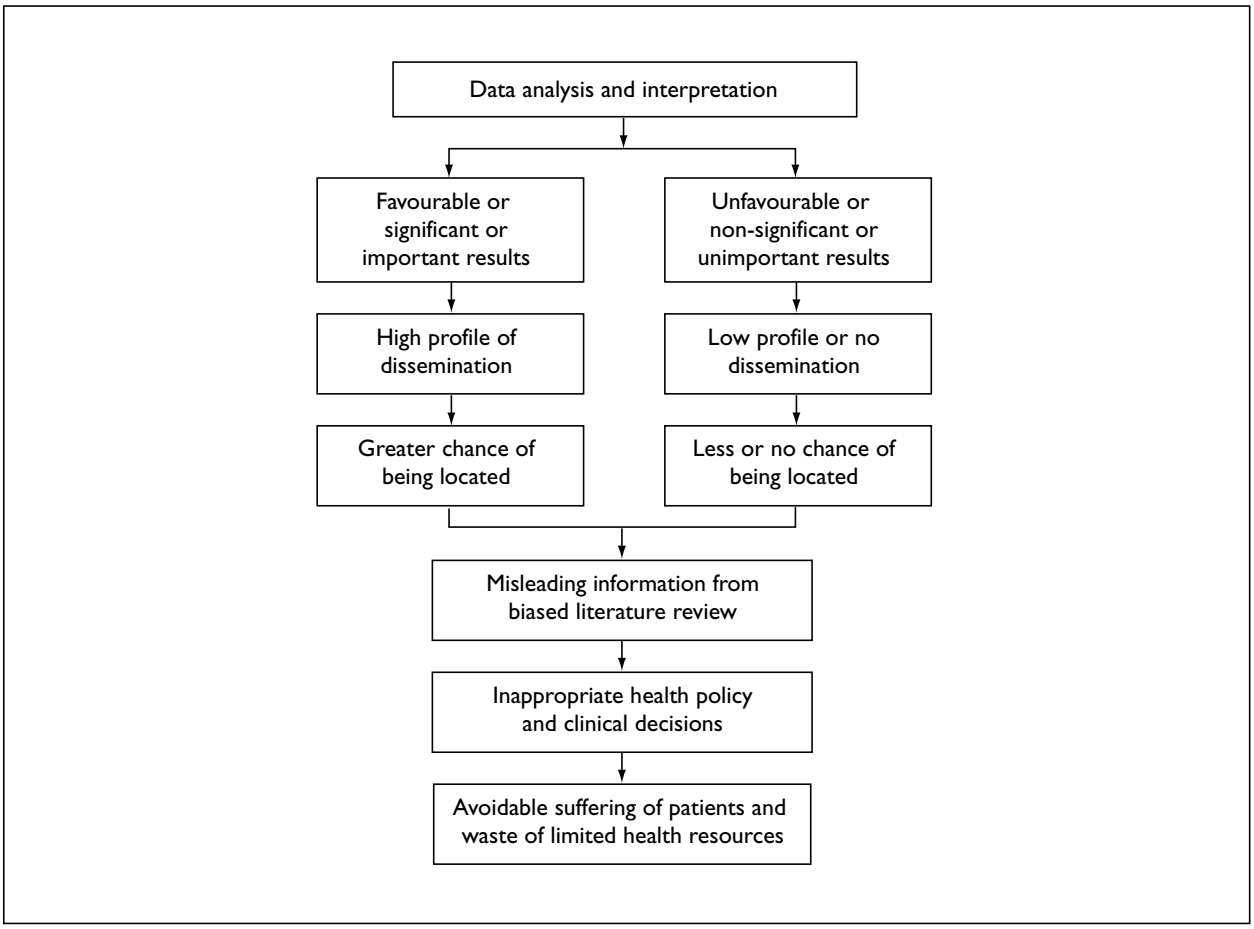


FIGURE 3 *The impact of publication and related biases*

there may be scepticism about new developments in health technology and controversies in medical research.

Patients' suffering and waste of resources

The impact of publication bias in clinical trials will depend on the direction and extent of the bias, and the underlying effects evaluated. The worst scenario is one in which a harmful intervention is falsely reported as effective because of publication bias. Patients may then receive a harmful treatment. If an ineffective intervention is falsely considered as effective, patients may receive ineffective treatment and be denied effective therapy. For an effective intervention, its effect may be overestimated because of publication bias. New interventions are generally more expensive than conventional interventions, so overestimation of the efficacy of new interventions is likely to result in increased cost without a corresponding improvement in outcome.

The non-publication of research findings may indirectly harm patients who are involved in future research. For example, a clinical study may find that an intervention is harmful but this finding is not published. Other investigators may subsequently repeat the same research, testing the harmful intervention on different patients.

A perinatal trial observed that routine hospitalisation was associated with more unwanted outcomes in women with uncomplicated twin pregnancies, but this finding remained unpublished for seven years.[143] Chalmers pointed out that "at the very least, this delay led to continued inappropriate deployment of limited resources; at worst, it may have resulted in the continued use of a harmful policy."[143]

In 1980, a trial tested lorcainide in patients with acute and recovering myocardial infarction. More deaths were observed in the treatment group than in the placebo group (9/48 versus 1/47).[144] The trial results were not published because the development of lorcainide was stopped for "commercial reasons". About a decade later, an increased mortality was observed among patients treated with encainide and flecainide in two trials.[145,146] Encainide, flecainide and lorcainide all belong to a class of I_C antiarrhythmic agents. If the results of the trial in 1980 had been published, the mortality of patients included in the two later trials might have been avoided or reduced. It was estimated that the number of extra deaths due to the use of class I antiarrhythmic drugs in the late 1980s was between 20,000 and 70,000 in the USA.[147]

Based on the results of a meta-analysis of several small trials, intravenous magnesium was recommended as a treatment for acute myocardial infarction.[148] However, a subsequent large trial (ISIS-4) showed that the death rate was higher in patients receiving intravenous magnesium than those receiving standard treatment (7.64% versus 7.24%; $p = 0.07$).[149] Publication bias has been identified as a possible explanation for the discrepant results from the meta-analysis and the subsequent large trial.[42] The number of hospital admissions due to acute myocardial infarction was 116,635 in 1993–1994 in England.[150] If the result of ISIS-4 trial was accurate and all these patients had received intravenous magnesium, there would have been 466 more deaths than if no patient had received it.

Controversies due to publication and related biases

Many controversies in the health and medical fields may be caused by the selective publication or citation of study findings.[7] Because of the early publication of supportive and delayed publication of unsupportive evidence for many new developments in health technology, "a wave of enthusiasm is sometimes followed by a wave of disappointment or scepticism".[57] One reason for the difficulty in changing professional behaviour may be that practitioners and researchers want to wait for confirmatory studies about a claimed advance before changing their practice. This scepticism about new developments may reduce the impact of publication bias, but the introduction of a new cost-effective technology may be unnecessarily delayed.

Changes in scientific thinking have led to the acceptance of multiple causes and uncertainty in explaining diseases. Many epidemiological studies have been conducted to identify and assess various risk factors.[151] However, there is contradictory advice about many risk factors identified by epidemiological studies.[152] For example, the results of epidemiological studies were contradictory about the risk of hair dyes, coffee, oat bran, oral contraceptives, environmental exposure to residential radon, and the presence of DDT metabolites in the bloodstream.[153] Publication bias may be an important

reason for many of the controversies surrounding the results of epidemiological studies.

Summary

Although the existence of publication bias is well demonstrated, there is little research on the impact of publication and related biases on health policy, clinical decision making and the outcome of patient management. The important consequences of publication bias include the avoidable suffering of patients and the waste of limited resources. At the very least, it is arguable that under-reporting research is scientific misconduct that may cause inappropriate patient care. In addition, it is unethical to abuse the trust of the patients involved and to waste invested resources.[16,143]

Chapter 5

Sources of publication bias

Publication bias may be introduced intentionally or unintentionally, consciously or unconsciously, because of varying motivations or biased standards used to judge research evidence.[154] Investigators, journal editors, journal peer reviewers and research sponsors may all be responsible for the existence of publication bias, although the extent of such responsibility may vary in different circumstances. In this chapter, available empirical evidence about sources of publication bias will be summarised. The results of computer simulations are then used to show that publication bias will be exacerbated by the great variation in the possible results from different studies.

Who is responsible for publication bias?

Investigators and authors

The reasons for not publishing completed research have been investigated in several studies (*Table 7*). These reported that investigators were the main source of publication bias, for not writing up or not submitting studies with null or unimportant

TABLE 7 *Reasons given by investigators for studies not being published*

Study	Reasons for non-publication[a]
Easterbrook *et al.* (1991)[51]	Total % of reasons given (*n* = 175): submitted for publication or published elsewhere 19%; null results 15%; limitations in methodology or logistic problems 12%; sponsor has control of data 11%; analysis incomplete 11%; manuscript rejected 9%; publication not aim of study 7%; too busy or lost interest 6%; unimportant results 6%; co-investigator left 3%
Dickersin *et al.* (1992)[52]	Manuscript rejected by journal 4.8% (6 of 124 unpublished studies) Main reasons for 118 not submitted: results not interesting 31.4%; design or operational problems 33.9%; publication not an aim 13.6%; other reasons 21.2%
Dickersin and Min (1993)[53]	Main reasons for 14 unpublished studies: not interesting or no time 42.8%; co-investigator or other operational problems 37.5%; data analysis not completed 14.3%; no reason 7.1%
Scherer *et al.* (1994)[68]	Unpublished abstracts of RCTs (*n* = 32): incomplete studies 15.6%; manuscript rejected 18.8%; no time to prepare 28.1%; problem of study design 9.4%
Rotton *et al.* (1995)[46]	Reasons given for not publishing: failure to replicate 4.8%; manuscript rejected 33.3%; non-hypothesised results 4.5%; inexplicable results 22.3%; non-significance 59.9%
Dickersin *et al.* (1987)[45]	Reasons for not submitting completed RCTs (*n* = 102): analysis in progress 14.7%; results negative 34.3%; lack of interest 15.7%; sample size or poor methodology 4.9%; controversy 2.9%; other or unknown 27.5%
DeBellefeuille *et al.* (1992)[69]	Reasons for non-publication of studies following submission of abstracts (total number of respondents *n* = 44): lack of time or other resources 29.5%; insufficient priority 20.5%; incomplete study 11.4%; manuscript rejected 9.1%; other 29.5%
Weber *et al.* (1998)[74]	Reasons for failure to submit a manuscript to a journal (total number of respondents *n* = 179): not enough time 41.3%; thought journals unlikely to accept 19.6%; results not important enough 11.7%; too much trouble with co-authors 8.9%; not worth the trouble 7.3%; other papers with similar findings 6.1%; statistical analysis not positive 3.9%; other reasons 22.3%
Misakian and Bero (1998)[58]	Reasons for unpublished results (*n* = 59) on effects of passive smoking: ongoing data collection or analysis 55.9%; lack of time 44.1%; competing priorities 18.6%; statistically non-significant results 3.4%; manuscript rejected 6.8%

[a]*There may be two or more reasons for each unpublished study*

results (range 18.6–42.8%). The unsubmitted studies as a proportion of all unpublished studies ranged from 47.6% to 100%. It was reported that quantitative studies with significant results were more likely to be submitted than studies with null results (78% versus 54%; $p < 0.001$). There was no difference in the publication rate of submitted manuscripts between studies with significant results and studies with null results (87% versus 82%; $p = 0.54$).[54] Studies with positive results were often submitted for publication more rapidly after their completion than were negative studies.[57]

In a case-control study[155] of 100 accepted and 100 rejected papers in two Spanish medical journals, it was found that publication status was associated only with high quality scores. Positive findings were reported in 143 of the 146 studies that used hypothesis testing. The author concluded that "if publication bias exists, it stems from authors, not editors."

Chalmers and colleagues[15] observed that the author's specialty was associated significantly with the enthusiasm for the procedure reviewed in an article. For example, 21 of 29 radiotherapists were enthusiastic for radiotherapy after radical mastectomy when stage is not distinguished, compared with only five of 34 authors from other specialties. In a systematic review of the risks of stroke and death due to endarterectomy for symptomatic carotid stenosis, it was observed that the risk was highest in studies in which patients were assessed by a neurologist after surgery and lowest in those having a single author affiliated to a department of surgery (7.7% versus 2.3%).[156] It is possible that surgeons were less likely to report the results if the operative risk of stroke and death was high.

A study evaluated authors' methods for selecting journals for the submission of manuscripts,[157] in which 64% of 479 questionnaires were returned and the factors that influenced selection were rated from unimportant (scale = 1) to very important (scale = 6). For initial manuscript submission, the mean value of importance was 5.2 for the journal's prestige, 4.8 for the makeup of the journal's readership, 4.8 for the usual topics of articles published in the journal, and 4.4 for the likelihood of manuscript acceptance. For a subsequent submission after a manuscript had been rejected by a journal, the most important factors were the likelihood of manuscript acceptance (5.0) and whether the journal usually publishes articles on the topic (4.7).[157]

To study the problem of data-withholding behaviours among academic life scientists, Blumenthal and co-workers[158] conducted a mailed survey of 3394 life science faculty members at the 50 universities that received the most funding from the NIH in 1993. Nineteen per cent of the 2167 respondents reported that publication of their research results had been delayed at least once by more than 6 months in the last 3 years. The reasons given for delayed publication were: to allow for patent application (46%), to protect their scientific lead (31%), to allow time to negotiate a patent (26%), to resolve disputes over the ownership of intellectual property (17%), or to slow the dissemination of undesired results (28%). In addition, 9% of the respondents reported that they had refused to share research results with other university scientists at least once in the last 3 years. Multivariate analysis revealed that delays in publication and refusal to share the research results were significantly associated with funding from industry or engagement in commercialisation.[158]

Editorial policies

Although editorial rejection was not a frequent reason given by investigators for studies remaining unpublished,[51,52] it cannot be ruled out that authors do not submit studies with negative results because of anticipated rejection according to journals' instructions to authors and their own experience. In a survey of 80 authors of articles published in psychology or education journals in 1988, 61% of the 68 respondents agreed that, if the research result is not statistically significant, there is little chance of the manuscript being published.[159] Weber and colleagues[74] found that anticipated rejection was given as a reason for failure to submit a manuscript by 20% of 179 authors. In another study, 17 of 45 submitted trials were rejected by at least one journal, and four negative trials with over 300 patients each were rejected two or three times, while no positive trial was multiply rejected.[57]

To study the impact of manuscript characteristics on journals' acceptance for publication, Kerr and co-workers[160] sent questionnaires during 1974 to 429 editors or members of advisory boards of 19 leading journals in management and the related social sciences. According to 301 responses, the likelihood of acceptance was increased by strong author reputation, successful testing of the author's own new theory, and content differing from that traditionally published by the journal. On the other hand, the following characteristics were considered to reduce the

chance of acceptance: non-significant results, replications, lack of new data, similar to recently published articles, or having previously been presented at meetings.[160]

The ideal articles for journals are those with findings that will affect clinical practice. It was suggested that negative results should be sent to a pay journal in order to be published because few people want to read the negative articles.[161] Sterling and colleagues[28] revealed a letter from an editor of a major environmental/ toxicological journal:

> "Unfortunately, we are not able to publish this manuscript. The manuscript is very well written and the study was well documented. Unfortunately, the negative results translate into a minimal contribution to the field."[28]

"Originality" is one of the most important criteria upon which journals decide whether a submitted paper will be accepted. It was suggested that the importance of a study could be decided according to whether its results are unoriginal, predictable, trivial, narrowly interested, highly specialised, or of few or no clinical implications.[162] According to *The Lancet*'s instruction to authors, articles published "are selected, from among a huge number of submissions, if they are likely to contribute to a change in clinical practice or in thinking about a disease."[163] Lack of originality accounted for 14% of all reasons given for the rejection of manuscripts in 1989 by the *American Journal of Surgery*.[164] Confirmatory trials, either positive or negative, have a low chance of being accepted.[165] A journal on diabetes clearly stated that "mere confirmation of known facts will be accepted only in exceptional cases; the same applies to reports of experiments and observations having no positive outcome."[166] The *NEJM* would normally reject epidemiological studies with a RR < 3.[153] Not surprisingly, editors will publish negative studies that may have a potential to change current practice by showing that a widely used intervention is ineffective.[167]

Peer reviewers

Journal peer review has been defined as "the assessment by experts (peers) of material submitted for publication in scientific and technical periodicals".[168] Godlee and Dickersin[169] classified unacceptable biases in peer review processes as biases in favour of or against certain types of author (prestigious or less prestigious, male or female, or those from particular countries) and biases in favour of or against certain types of manuscript (with innovative ideas, written in languages other than English, with positive or negative results).

The degree to which peer review contributes to publication bias was investigated by Mahoney, who examined the recommendations of 75 journal referees about a fictitious manuscript with identical experimental procedures but different results.[170] It was found that inter-rater agreement was poor and that referees were biased against the manuscript that reported results contrary to their own perspectives (confirmatory bias).

Ernst and Resch[171] sent a fictitious research paper with only study design and results to 33 referees. The topic of the fictitious research was about transcutaneous electrical nerve stimulation (TENS) and referees were identified by their previous publications as pro-TENS or contra-TENS. As in Mahoney's study, it was found that inter-rater reliability was poor, and referees' judgement was associated with their own preconceptions and experience.[171] In another study, Epstein sent two versions of a fictitious paper to 146 social work journals, one with a positive result and one with a negative result.[172] The manuscript was accepted by six of the 17 journals that reviewed the positive version and four of the 16 journals that reviewed the negative version (35% versus 25%; $p > 0.05$).[172]

To study whether referees in complementary medicine are biased, different versions of the same fictitious short communication were sent randomly to 200 authors of articles that were selected from MEDLINE for the period 1994–1996.[173] The same manuscript was artificially designed to fit the following four versions: good quality with a positive result, good quality with a negative result, poor quality with a positive result, or poor quality with a negative result. It was found that the poor quality manuscript was more likely to be rejected than the good quality manuscript (55% versus 16%; $p < 0.05$), and there was no evidence of reviewing bias against a positive or negative outcome.

Conclusions were different from two studies that evaluated peer-reviewer bias against unconventional treatment.[174,175] One study found that peer reviewers showed a bias against papers dealing with unconventional medical concepts.[174] For two versions of a fake report on obesity treatment, reviewers' rating of the orthodox version was on average significantly higher than that of the homoeopathic therapy version. However, reviewer bias against the unconventional drug was not observed in a study

that used two versions of a fake report about an in-vitro experiment of a mainstream drug or a highly unconventional yet commercially available drug.[175]

One hundred and eighty Scandinavian referees were asked to review two fictitious but realistic short manuscripts with some methodological flaws.[109] They all received a manuscript in English and another manuscript in their national language. A total of 156 referees returned 312 reviews. The quality of the English version was considered to be better than that of the national-language version ($p < 0.05$). It was concluded that "an English version seemed to be accepted more easily than a national-language version of the same manuscript."

In a retrospective study, the effect of institutional prestige on referees' recommendations and editorial decisions was assessed.[176] Institutional prestige was determined according to the monetary value of research and training grants and contracts funded by the NIH. The association between the recommendation for acceptance and institutional prestige was observed for the 147 brief reports (i.e. case reports and similar short papers) but not for 258 major papers (such as case series, research reports, epidemiological studies).

Ector and colleagues[177] compared the agreement between reviewers in grading abstracts submitted to a conference (the Sixth European Symposium on Cardiac Pacing). Each abstract was graded on a scale of 1–10 by two peer reviewers. There was no statistically significant correlation between reviewers in 13 of the 28 pairs. It was suggested that reviewing abstracts is less predictable and more likely to be biased than reviewing a full article.

Readers and users of research findings

Journal editors' policy may reflect readers' preferences, and it has been suggested that editors should find ways to incorporate the reader's perspective into the peer review process and study the effects of their efforts.[178] It is likely that readers' preferences for certain findings may be an important reason for the biased publication of studies in journals.

In a survey of 452 readers who indicated that they were likely or highly likely to read a manuscript, Justice and co-workers[178] reported that readers were generally satisfied with the quality of manuscripts but relatively dissatisfied because of a lack of manuscripts that are relevant to medical practice. Differences of opinion between readers and peer reviewers may be due to the fact that clinicians tend to avoid unestablished treatments but research articles with new information are more likely to be accepted for publication.[178]

Research funding bodies and commercial interests

According to a telephone survey of 306 companies in 1994 (response rate 69%), Blumenthal and colleagues[179] found that 59% of companies conducting life science research in the USA supported research in an academic institution, providing approximately 11.7% of all the research and development (R&D) funding received. When compared with a similar study carried out in 1984,[180] it was found that life science companies were more likely to support academic research in 1994 than in 1984 (57% versus 46%; $p = 0.05$). Over 60% of the companies providing support for life science research in universities had received patents, products and sales as a result of these relationships.[179] Eighty-two per cent of the companies surveyed asked investigators at universities "to keep the results of research secret beyond the time needed to file a patent".[179]

Blumenthal and co-workers[181] also surveyed 2052 life sciences faculty members at 50 universities in the USA by means of a mailed questionnaire (response rate 65%). It was found that 28% of the respondents received research support from industry, and that the proportion of industrial funds was greater for clinical faculty members than for non-clinical members (12% versus 6%). Faculty members with industrial support were more likely to report restriction on the dissemination of their research findings (15% versus 5%) because of "trade secrets" that were defined as information kept secret to protect its proprietary value.[181]

Rosenberg, according to his own experience, suggested that secrecy in science "has escalated dramatically in the past decade and is impeding the progress of medical research".[182] Because of the increasing support of medical research by biotechnology and pharmaceutical companies, new sources of funding are provided but increasingly "scientists are pulled in opposite directions by the desire to share research findings and the need to protect the investors who have supported the research". In Rosenberg's article, the major concern seems to be the withholding of information on new developments that are shown to be efficacious.

Many clinical trials submitted by drug companies to licensing authorities have never been published. Thirty-eight per cent of trials submitted by drug companies to licensing authorities in Finland or Sweden in 1974 and 1975 were not published. The quality of the unpublished studies was similar to those that were published.[183] In Finland, 274 notifications of the commencement of a clinical drug trial were received in 1983 by the National Agency for Medicine.[184] By the end of 1993, 68 of these 274 trials reported their results; 24 were suspended and the sponsors of the remainder were requested to report the outcome. It was found that the rate of reporting was 38% for trials with positive results, 18% for those with inconclusive results, and 20% for those with a negative outcome ($p = 0.023$).[185]

Commercial interests and intellectual property are often the main reasons for the non-publication of clinical studies funded by drug companies.[186–188] As quoted in an editorial in *JAMA*, in one study it was reported that 35% of the signed agreements in a sample of university–industry research centres allowed the sponsor to delete information from the publication, 53% allowed publication to be delayed, and 30% allowed both.[189] Abraham and Lewis[187] argued that:

> "the present European medicines licensing system is biased in favour of commercial interests at the expense of medical science, public health, and the adequate provision of information for doctors and patients."

Drug companies may be particularly unwilling to publish sponsored trials with unfavourable results.[190] Rennie has described an example in which a pharmaceutical company tried extremely hard to block publication of the negative results of a study it had sponsored.[189] A study carried out by Dong and co-workers[191] evaluated the bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. The price of the company's product could have been affected by the study's conclusion that the generic and brand-name levothyroxine preparations are bioequivalent. The study was completed by 1990 and the manuscript was accepted for publication by *JAMA* in 1994. However, the manuscript was not published until 1997 because the pharmaceutical company "waged an energetic campaign to discredit the study", threatened to carry out legal action against the investigators and the investigators' university.

Companies may also try to prevent the publication of studies conducted by others when the findings will undermine their commercial interests. For example, a pharmaceutical company attempted to prevent the publication of a systematic review that would have a negative economic impact on statins (cholesterol-lowering drugs).[192] In another case, a company that produces hormone bovine somatotropin blocked the publication of a meta-analysis with unsupportive results by using its legal rights over the raw data.[193]

Clinical trials published in 1984 in five general medical journals were selected to compare funding source and the trial outcome.[33] Drug company funded trials were significantly more likely than generally funded trials to support a new therapy ($p = 0.002$); the proportion of trials favouring a new therapy was 89% in 37 trials funded by drug companies and 61% in 70 trials funded by other sources. It was suggested that this finding might be related to many factors such as publication bias, the selection of drugs likely to be proved efficacious, false-negative results, and fear of discontinuation of funding.[33]

Rochon and colleagues[194] searched MEDLINE to identify RCTs concerning NSAIDs in the treatment of arthritis that had been published between September 1987 and May 1990. The manufacturer-associated drug was defined as the drug of interest to the sponsoring company and was identified on the basis of information provided in the article or from standard references. The manufacturer-associated drugs were reported to be comparable with (71%) or superior to (29%) the control drugs in all 56 trials. These narrative claims of superiority were usually justified with trial data. Of the 22 trials that reported a drug with less toxicity, the manufacturer-associated drug's safety was reported to be superior in 86% of cases. Justification for the narrative interpretation of the trial findings concerning less toxicity was provided in only 12 (55%) of the 22 trials. It was concluded that "these data raise concerns about selective publication or biased interpretation of results in manufacturer-associated trials."[194]

Stelfox and colleagues[195] examined the relationship between authors' published positions on the safety of calcium-channel antagonists and their financial interactions with the pharmaceutical industry. They identified 77 articles (including reports of original research, reviews and letters to the editor). A questionnaire was sent to 86 authors of the 70 articles to ask about their financial interactions with pharmaceutical companies. A total of 69 authors completed

the survey. It was found that 96% of the supportive authors had financial relationships with manufacturers of calcium-channel antagonists compared with 60% of the neutral authors and 37% of the critical authors ($p < 0.001$). Although there are limitations in Stelfox and co-workers' study,[195] as mentioned in the discussion section by these authors and in four letters to the editor,[196–199] the importance of full disclosure of relationships with pharmaceutical manufacturers in journal articles is highlighted.[200]

Barnes and Bero[201] examined research sponsored by the tobacco industry through the Centre for Indoor Air Research. This Centre's special-reviewed projects are more likely than its peer-reviewed projects to support the tobacco industry position and to be used by the industry to argue that smoking should not be regulated in public places. Symposium articles were also more likely to agree with the tobacco industry's position and more likely to be written by tobacco industry-affiliated authors than were journal articles.[91]

Cho and Bero[202] compared drug studies published in symposium proceedings that are sponsored by drug companies with articles published in their parent medical journals. There was no significant difference in the mean methodological and relevance scores between the two groups. The percentage of articles in favour of the drug of interest was 98% for articles with drug company support and 79% for those without drug company support ($p = 0.01$). The potential reasons for this phenomenon were discussed. For example, drug companies will be more willing to support positive findings and the studies sponsored by the drug companies are more likely to use placebo controls in order to prove the drug's efficacy.[202]

Statistical considerations

If the results from all possible studies were the same or similar, it would not be possible to select particular results for publication. Greater variation in the results of studies may be associated with an increased risk of publication bias. Factors that influence variation in study results include small sample size, small or moderate effect size, subjective nature of the outcome measurement, and complex interventions.

Small sample size

Studies with small sample sizes tend to produce very variable results and present a range of results to select for publication. Simulations have demonstrated that small sample size is associated with considerable publication bias when only studies with significant results are published.[203,204]

In practice, a small study with a non-significant result may be readily abandoned without trying to publish it because it is easy and cheap to carry out in terms of time, staff and other resources invested. In addition, small trials may often be poorly designed and conducted. Therefore, the risk of publication bias will be great if many small trials have been conducted.[205]

Figure 4 shows the results of a stochastic simulation investigating the relationship between publication bias and the range of possible sample sizes. Given a true OR of 0.73 and other conditions assumed in the simulation, the estimated OR is 0.23 (corresponding to an absolute difference in lnOR of 1.135) when the sample sizes range from 20 to 100, 0.44 when they range from 20 to 500, and 0.70 when the range is from 20 to 5000. When the possible sample sizes range from 20 to 10,000, the estimated OR is 0.72, which is nearly identical to the true value of 0.73. That is, when there are many large-scale trials, the extent of bias due to the selective publication of significant results will reduce.

Small or moderate effect size

The simulation results also indicated that, when only significant results are selected for publication, the extent of bias is greater when the true effect is small or moderate than when the true effect is zero or very large.[203] *Figure 5* shows the results of a computer simulation investigating the relationship between the true effect (OR) and the extent of bias. The extent of bias is small when there is no effect at all (OR = 1.0), and increases when there is a small effect. Under the given simulation conditions, the difference between the true and the biased effect was largest when the true OR is about 0.85. When the assumed true treatment effect increases further (that is, OR becomes further away from 1.0), the simulated bias becomes smaller. However, the bias is still considerable (ratio of ORs = 0.80) even when the true OR equals 0.40. Therefore, a small or moderate effect (or weak association) can be considered as a risk factor for publication bias. This risk factor may exist in most cases because clinical trials are mainly designed to assess health care interventions with small or moderate (but clinically important) effects.

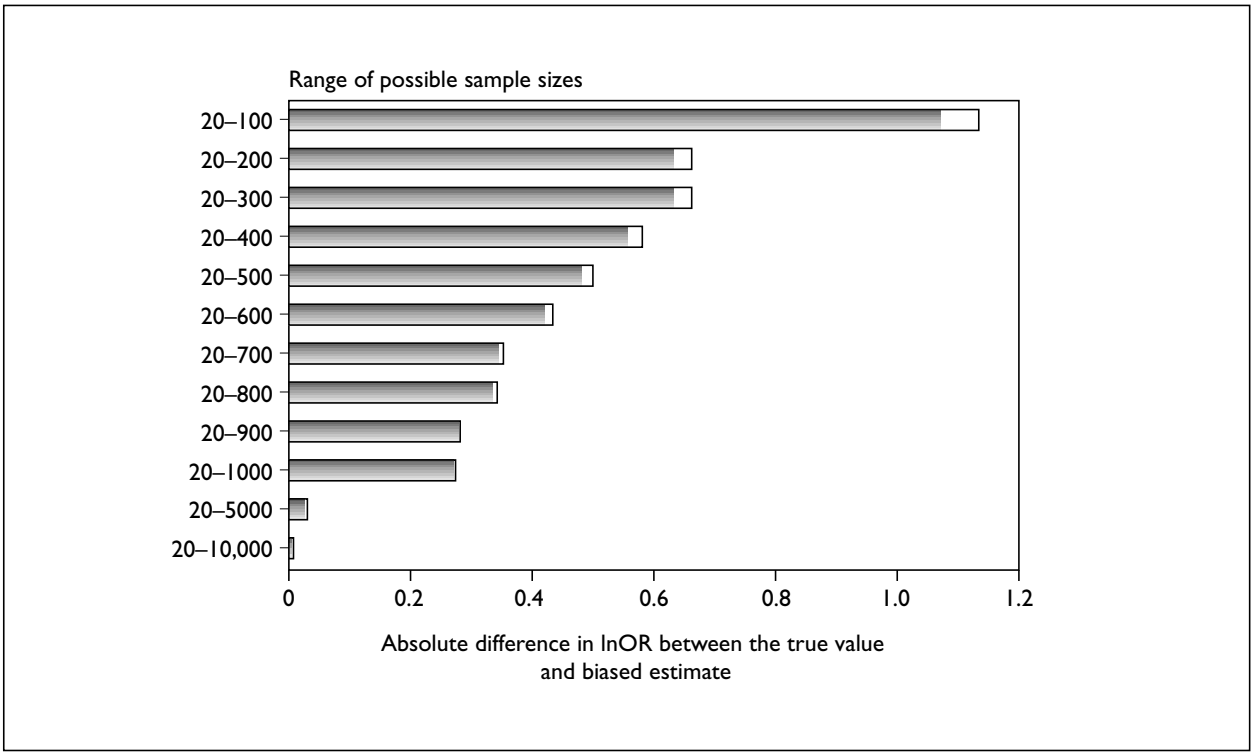


FIGURE 4 *Extent of publication bias and range of possible sample sizes: results of a stochastic simulation* (trials conducted n = 500, selected when |Z| > 1.96; rate in control group 0.1; true OR 0.73 (natural log OR (lnOR) – 0.315); see appendix 4 for more details)

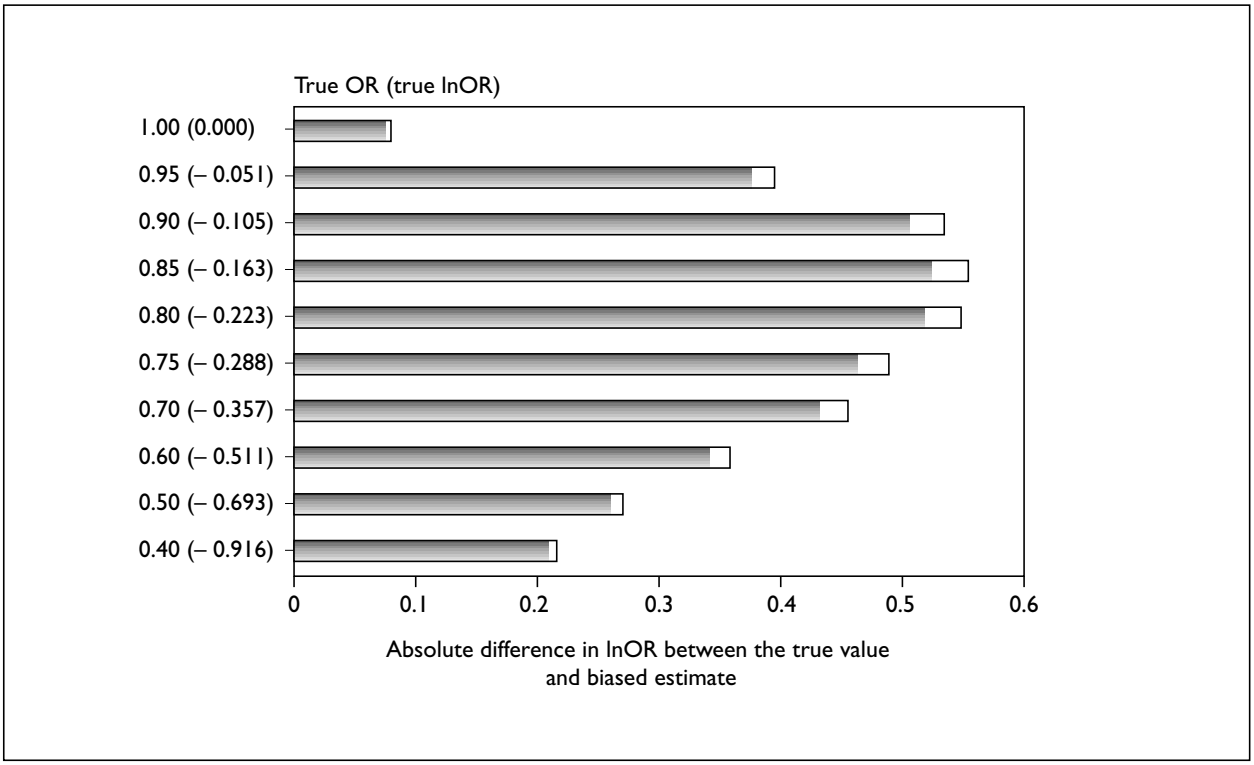


FIGURE 5 *Extent of bias and the true effect size (lnOR): results of stochastic simulations* (trials conducted n = 500; selected when |Z| > 1.96; range of possible sample sizes 20–500; event rate in control group 0.1; see appendix 4 for more details)

Study design and other quality characteristics

The design quality of studies may be associated with the risk of publication bias. Non-randomised studies, single-centre studies, and Phase I and II trials may be more susceptible to publication bias than randomised studies, multicentre studies, and Phase III trials.[39,206] Risk factors for publication bias were assessed but not consistently identified across several cohort studies of publication bias.[51–54] Irwig and colleagues[207] suggested that publication bias is more of a problem for diagnostic tests than for randomised trials because "many studies of test accuracy may use data collected primarily as part of clinical care, [and] there may be no clear record of attempted evaluations".

Summary

Investigators, peer reviewers, editors and funding bodies may all be responsible for the existence of publication bias. The importance of users of research results and readers of journals cannot be ignored. Considering the strong motivation for investigators to publish, it is surprising that publication bias is often due to their failure to write up their research and submit it for publication. However, it should be recognised that the decision to write up an article and then submit it may be affected not only by their own preferences but also by pressure from research sponsors and instructions from journal editors. Evidence shows that the interest of research sponsors can restrict the dissemination of the research findings. It may also be true that a journal's interests affects the Editor's decision concerning acceptance or rejection of a manuscript. The dissemination profile of a research finding is determined by the interests of research sponsors, investigators, peer reviewers and editors. The large potential variation in results obtained across similar studies that can easily be conducted and abandoned will further exacerbate the biased selection of findings for publication.

Chapter 6

Prevention of publication bias

Biased selection may occur at a number of stages in the publication process: when a manuscript is prepared and then submitted; when a submitted manuscript is peer reviewed, and when the editor makes a decision regarding which articles to include. Measures to prevent publication bias should logically be designed according to the sources of the bias. Although investigators, peer reviewers, editors and funding bodies may all be responsible for the existence of publication bias, the importance of their responsibility in terms of preventing it may vary. For example, the curiosity of general readers about new and unusual events is very difficult, if not impossible, to change. It is understandable and it may also be necessary for readers to select relevant journal articles. Journal editors who want to increase or maintain their circulation level will then have to select studies for publication according to their perception of their readers' preferences and the type of information those readers require.

In spite of these difficulties, it is possible that the biased publication of research, or its impact, may be prevented. In this chapter, changes in journals and the publication process that could prevent publication bias are discussed. Then it is proposed that prospective registration of studies at their inception is the best preventive solution. Finally, it is suggested that the impact of publication bias may be prevented by conducting confirmatory large-scale trials.

Changes in editorial policy and the publication process

Medical journals have two basic functions: of medical recorder and medical newspaper.[208] The function of a medical recorder is important in order to facilitate communication between investigators for the advancement of medical knowledge. The function of a medical newspaper is to disseminate information that is relevant to medical practice. It seems that conventional paper journals have many intrinsic limitations as medical newspapers and medical records. The major concern in this review is about their biased publication of research findings.

It has been suggested that journals could help to reduce publication bias by accepting manuscripts for publication that are based mainly on the research protocol.[209] To motivate investigators to register their trials, Julian[210] suggests that prospective registration should be a requirement laid down by the editors of journals and registering agencies. By disclosing a "conflict of interest" or "competing interests", potential bias due to sources of research funding may be revealed.[200]

Recently, over 100 medical journals around the world have invited readers to send in information on unpublished trials in a so called "trial amnesty".[211,212] By the end of 1998, this amnesty had registered 150 trials.[213] Since the beginning of 1997, a general medical journal, *The Lancet*, has been assessing and registering selected protocols of randomised trials and providing a commitment to publish the main clinical findings of these studies.[163] However, it is unlikely that traditional paper journals can solve with significant impact the problem of the selective publication of studies, because of space limitation and the requirement of "newsworthy" articles for maintaining or increasing the journals' circulation levels.

A considerable proportion of trials will not be able to provide results that are statistically significant or clinically important if each of these trials is considered separately. The impact of many trials on clinical practice may not be immediately clear and they may have a very limited readership. Therefore, conventional medical journals are not enthusiastic about findings that are seemingly of no impact or of little interest. However, the findings of these trials are very important for the unbiased evaluation of health care interventions.

The recent emphasis on evidence-based medicine may stimulate more clinical trials to be carried out.[214] However, in four general medical journals, the number of trials reported reduced by about 50% from 1980s to 1990s.[215] The reasons why general medical journals now publish fewer clinical trials than previously may be due to reduced submissions of trial findings, or to increased rejections of trial findings that have been submitted. If many trial findings are rejected after peer reviewing

because of a lack of space and/or a lack of importance, it will be a waste of resources. Authors of a manuscript that is rejected by one journal will want to submit it to another journal. The editors and peer reviewers of another journal will then have to repeat the editorial and peer reviewing process. Therefore, publishing studies with non-significant results may be more expensive and take longer than publishing studies with significant or striking results.

Peer review process

Peer review may improve the general quality of published studies.[216] One study showed that 62% of manuscripts rejected by the *American Journal of Surgery* were not published in other indexed medical journals, and it was concluded that "peer review is an effective screening process to evaluate medical manuscripts".[164] One hundred readers of the *Nederlands Tijdschrift voor Geneeskunde* (*Dutch Medical Journal*) were invited to evaluate the submitted, accepted and published versions of 50 articles.[217] The quality score and general medical value of manuscripts were significantly improved after peer review ($p < 0.01$). However, publication bias may occur because of biased peer review. Therefore, the selection of peer reviewers should be balanced, particularly when the topics reviewed are controversial. Recently, more studies have been conducted to examine the peer review process and the move towards evidence-based peer review may help to prevent or reduce bias arising at this stage of the process.[218]

Although the results from different studies were conflicting, it seems that the general quality of peer review has not been improved by blinding peer reviewers to authors' identities.[219–221] In fact, it is often impossible to have a fully closed peer review (neither the reviewers nor the editors know the names or institutions of the authors, and the authors do not know the names of the reviewers).[222–225] Thus, Rennie argued that a full open system of peer review will strengthen "the link between power and accountability, because when reviewers know their names will appear at the end of their reviews, one may be sure that they will be constructive and will attempt to back up their statements."[226] Recently, the *BMJ* started an open peer review system by asking reviewers to sign their reviews so that authors would know who had assessed their articles.[227]

Electronic journals

Along with the development of computer science and internet technology, there has been a rapid increase in the volume of electronic publishing. Compared with paper journals, electronic journals have many advantages, including timely publication, no limitation on the length of articles or number of studies, linkage of an article to other studies or comments, and being relatively cheap to disseminate and preserve.[228–232] Although a conventional paper journal may be a biased "medical recorder", both of the two basic functions of journals may benefit from electronic publishing. It has been suggested that, in the electronic era, publication is becoming more database-like and, at the same time, databases are starting to become more like publications.[233]

Sim and Rennels suggest that trials should be published concurrently in electronic database form and traditional prose form according to a Trial Bank Model.[234] However, it seems that the Trial Bank Model aims to provide extra data from trials that have been published in paper journals rather than to prevent publication bias. Recently, some journals have introduced electronic supplements that are used for publishing additional information relevant to the articles appearing in the paper version. For example, the *BMJ* is experimenting with so-called ELPS (electronic long, paper short), by publishing a short version of some studies in the paper version and a longer version in the electronic form.[235]

Chalmers and Altman[236] suggest that electronic publishing provides opportunities for improving the quality of medical research by publishing research protocols in advance of publication of the full study or even prior to its commencement. By November 1998, nine trial protocols had been listed on *The Lancet*'s website[163] in response to this journal's initiative as already noted.

The potential of electronic publishing in reducing publication bias has been recognised.[143,236–238] For example, Berlin states that papers submitted to the *Online Journal of Current Clinical Trials* will be judged "on the quality of the work and the importance of the findings, not on the level of statistical significance achieved in any comparisons made."[237] However, publication bias may still occur if a study's perceived importance is associated with the direction or magnitude of the findings.

Publication bias could be reduced through introducing new peer reviewed electronic journals that have no space limitations and in which the basis of manuscript acceptance is on the research

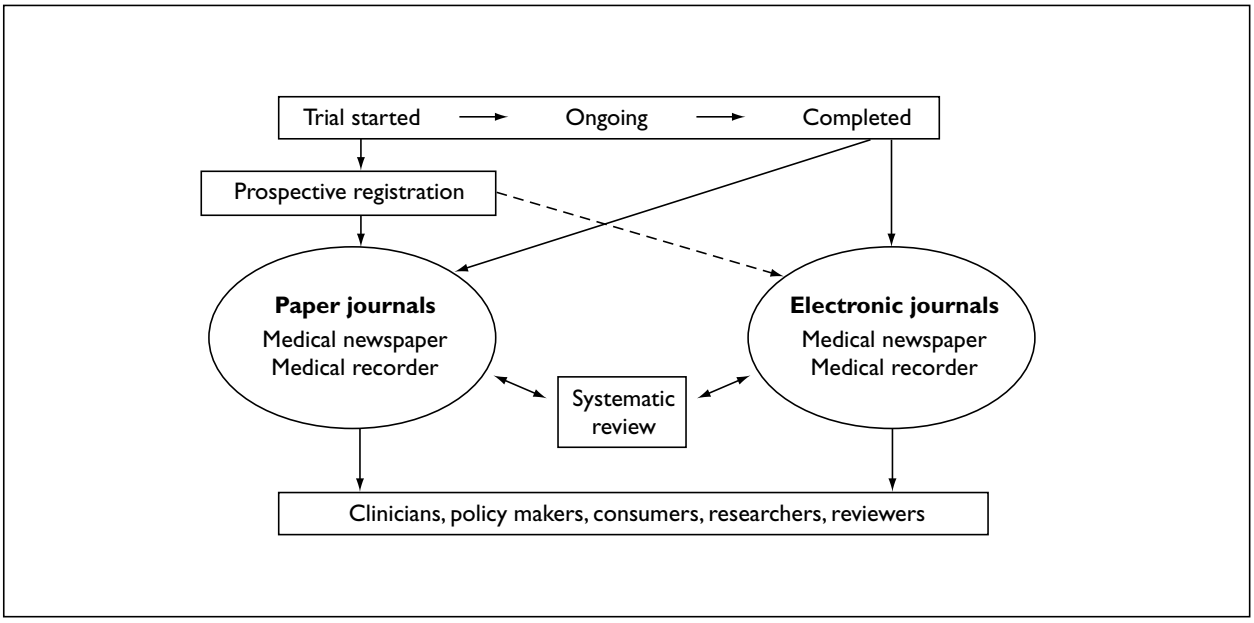


FIGURE 6 *Electronic journals for reducing publication bias (redrawn from Song* et al. *(1999)[239])*

methodology or validity criteria only, rather than on the findings (positive or negative, important or not important, significant or non-significant).[239] If originality or newsworthiness were to become less of a requirement for publication, then this type of electronic journal would encourage the submission of trials with negative or non-striking results, and also of trials that replicate previous studies. Such electronic journals may be used as medical recorders, to document and store findings from all clinical trials, including those that seem to be of no immediate impact on practice. Conversely, paper journals could focus more on the findings that have an important impact on clinical practice or are of wide professional interest (*Figure 6*).

Although the incentive to publish is strong, authors may not wish to spend much time preparing manuscripts that have little chance of being accepted by peer reviewed journals. Similarly, they may have limited motivation to publish their research in a databank or "trial amnesty" without receiving credit from such publication.[238] Because it is generally believed that peer reviewed journals are of better quality,[162,217] electronic journals publishing trials should be peer reviewed. Authors will be more likely to submit their research to peer reviewed electronic journals that have a good reputation for quality. As in conventional journals, the main task of peer reviewing is to assess the quality of submitted studies and to maintain the academic rigour of a journal. Accepted studies may be published together with peer reviewers' comments about their methodological strengths and limitations.

The establishment of unbiased electronic journals of clinical trials may not be an easy task. An early electronic trials journal, the *Online Journal of Current Clinical Trials*, ceased to exist for unknown reasons (Berlin J, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Health System: personal communication, 1998). However, more electronic journals have been or will be established. Recently, the Current Science Group has announced it is to establish "a series of peer-reviewed controlled trial journals featuring protocols, reports on trial design, results, and trial data (with no restriction on space)."[240] It has been clearly stated that "selection criteria will be based purely on the methodological quality of the trial and not on the newsworthiness of its results."[240]

Prospective registration of trials

Boissel and colleagues[241] defined a clinical trial registry "as a database of planned, ongoing or completed clinical trials, published as well as unpublished, in which details concerning the trial's objectives, main design features, sample size, and tested treatment are stored." According to several articles,[242–244] the main aims of the trial registries are:

- to facilitate communication and cooperation between researchers
- to help in planning future research and preventing unnecessary duplications

- to publicise trials and increase patient accrual to trials
- to help funding bodies in assessing new proposals for trials
- to prevent publication bias in systematic reviews and meta-analyses.

There is little disagreement that prospective registration of all trials at their inception is the best solution to prevent publication bias.[11] If it is not possible to register all trials, the prospective registration of some trials may provide an unbiased sample of all studies that have been conducted.[7] For example, the International Cancer Research Data Bank was used to assess alkylating agent therapy in advanced ovarian cancer.[56] The Stroke Unit Trialists' Collaboration[245] compared the results of published trials with the results of trials that were prospectively identified before they were published or before their data were analysed. However, it has been argued that assessment even in this case may be biased in favour of priors in a Bayesian analysis because the incomplete use of empirical evidence may add unwarranted weight to the priors based on subjective judgement.[246]

The Clinical Trials Registry of the International Committee on Thrombosis and Haemostasis may be the first registry of clinical trials.[247] Easterbrook surveyed 62 organisations and 51 individuals involved between March and June 1989 in clinical trials in 13 countries.[243] Twenty-four registries of clinical trials were identified according to 82 respondents. The registries were often disease specific and covered the areas of oncology ($n = 8$), AIDS ($n = 7$), cardiovascular disease ($n = 2$), neurosurgery ($n = 1$), rheumatology ($n = 1$), dentistry ($n = 1$) and perinatal medicine ($n = 1$). The survey also identified three multidisciplinary registries: the Spanish Database of Clinical Trials, the NIH Inventory of Clinical Trials, and the Department of Veterans Affairs Register of Clinical Studies.[243]

Since the publication of Easterbrook's directory, more registries of clinical trials have been established (see appendix 5).[122,213,248–253] The National Research Register is a register of ongoing and recently completed research projects funded by, or of interest to, the National Health Service in the UK.[251] Issue 2 of the National Research Register contains information on over 42,000 research projects, as well as entries from the Medical Research Council's Clinical Trials Register and details on reviews in progress collected by the NHS Centre for Reviews and Dissemination.[251] Another major development is the establishment of the Internet-based metaRegister of Controlled Trials by Current Controlled Trials Ltd.[254] This aims to "provide free access to a comprehensive database of ongoing and completed trials in all areas of healthcare", and had included 58 registers of trials by the end of 1999.

In Easterbrook's survey,[243] it was found that registers identify clinical trials: by surveying selected individuals, organisations, pharmaceutical companies or other industries; from conferences and selected journals; by searching other related registries of trials; and by funding bodies or RECs. It should be stressed that many published trials have been identified retrospectively and included in registries. For example, the trials included in the Cochrane Controlled Trials Register (CCTR) were identified by handsearching journals, conference proceedings and other sources. It includes many trials in "grey literature", but the potential for publication bias cannot be ruled out.

RECs may play an important role in eliminating publication bias by requiring the registration of trials at their inception and also a commitment to disseminate the research results as a condition of approval.[255,256] In Spain, a register of clinical trials has been established as a consequence of the law: a Royal Decree of 1978 and a Ministerial Order of 1982.[257] In the USA, the FDA Modernisation Act of 1997 includes a section that calls for the establishment of a federally funded database containing information on both government-funded and privately-funded clinical trials of drugs designed to treat serious or life-threatening conditions.[258]

Some pharmaceutical companies may now be willing to register in publicly accessible databases the trials they sponsor. For example, Glaxo Wellcome Ltd will provide a comprehensive record of all Phase II, III and IV studies conducted on newly registered medicines.[254] Another drug company, Schering Healthcare Ltd, has agreed to provide data on Phase III trials conducted over the past 5 years (published or unpublished) for reviews conducted by the UK Cochrane Centre.

It was hoped "that prospective registration will be the norm for all clinical trials by the time we enter the twenty first century."[255] However, there are many difficulties in establishing and maintaining registries of clinical trials.[259] In addition to a lack of funding,[260]

some pharmaceutical companies may refuse to provide information on trials in progress for various reasons, in particular because their competitive advantage may be reduced by a loss in confidentiality.[261] Other problems include the lack of incentives for researchers to register trials and competition among organisations that are operating their own registers.[259]

However, trial registers have been developing rapidly and becoming an increasingly important source of research information. Dickersin and colleagues identified less than 100 registers of clinical trials (mostly available on paper only) from 1987 to 1998. Two years later, the number of identified online registers of clinical trials exceeded 500.[262]

Freedom of information

Since 1997, the practice of the incomplete release of information in Europe about licensed drugs for reasons of commercial interest and intellectual property has been challenged.[186–188,263,264] Abraham and Lewis[187] suggested that "the secrecy and confidentiality of EU medicines regulation is not essential for a viable pharmaceutical industry", considering that European pharmaceutical companies often obtain data on competitors' products by using the US Freedom of Information Act. Confidentiality is no longer an important issue in registering late Phase II and Phase III clinical trials in which many hospital staff and hundreds or thousands of patients have already been involved.[258] It was found that the majority of industrialists and regulators interviewed (74%) did not, in principle, oppose greater public access to information.[187]

Rennie recommended that "investigators should not assume that the sponsors will encourage publication of unfavourable results and should never allow sponsors veto power."[189] Rosenberg suggested that scientists should refuse "to keep information confidential and refuse to sign any agreements for the transfer of information or reagent that included a requirement of confidentiality".[182]

There are some "encouraging signs" within the pharmaceutical industry that public access to the findings of industry-sponsored clinical studies has improved.[188] One pharmaceutical company changed its policy about dissemination of the research it sponsors, allowing investigators:

> "to publish studies conducted under generally accepted standards of scientific rigour without company prior approval, [subject to the] right to review prepublication drafts to address intellectual property issues, ie, matters involving patents, copyrights, and related trade secret information such as process and formulation information."[265]

Confirmatory large-scale trials

For the purpose of avoiding moderate biases and moderate random errors in assessing or refuting moderate benefits, large numbers of patients are required in RCTs.[266] Large-scale trials are generally believed to be less vulnerable to publication bias; this is the fundamental assumption of many methods for its detection. When the existence of publication bias is likely and the consequence of such bias is clinically important, a confirmatory, multicentre large-scale trial may be conducted to provide more convincing evidence.

It has been suggested that there should be more comparisons of meta-analyses of small studies with a "gold standard" (large cooperative trials).[267] Villar and co-workers[268] removed the largest trials from 30 meta-analyses in the Cochrane systematic reviews for pregnancy and childbirth. They then compared the results of the largest trials with those of meta-analyses after excluding the largest trials. It was found that 24 meta-analyses correctly predicted the direction of the treatment effect, but only 18 of the 30 were the same in both direction of the treatment effect and the statistical significance as the largest trial.[268]

From the Cochrane Pregnancy and Childbirth Database and from MEDLINE (1966–1995), Cappelleri and colleagues[269] identified 79 meta-analyses containing at least one large study of 1000 or more patients, and 61 meta-analyses with at least one large study based on statistical power considerations. They found agreement between large and smaller trials in 90% of the meta-analyses selected by the sample size approach and in 82% of the meta-analyses selected by the statistical power approach, by using a random effects model. Plausible explanations were identified in ten of the 15 disagreements between the results of large and small trials. The disagreements could be explained by differences in the control rate of events ($n = 5$), specific protocol or study differences ($n = 4$), and potential publication bias ($n = 1$).[269]

A different approach was used by LeLorier and co-workers[270] to identify meta-analyses and subsequent large trials. They first identified large trials that included 1000 or more patients, which were published in four general medical journals. They then identified meta-analyses that had been published earlier on the same topics. They concluded that the outcome of the 12 large trials "were not predicted accurately 35 percent of the time by the meta-analyses published previously on the same topics."[270]

The results of these three studies that compared large trials and meta-analyses of small trials should be interpreted with caution. For example, LeLorier and colleagues' study was criticised for the simplistic definition of positive and negative trials, and also for oversight of some important large trials or meta-analyses.[238] By evaluating the protocols of three studies that compared meta-analyses and large trials,[268–270] it was found that the conclusions of studies comparing large trials and meta-analyses may vary because of different methods for selecting trials and meta-analyses, and different methods for defining end-points and agreement.[271]

Well-designed small trials could be as reliable and valid as large-scale trials if the small trials could be prospectively registered at their inception. In addition, the systematic review of small studies may provide useful information about whether a confirmatory large study is required and how to design such a study.[272] Large-scale confirmatory trials become necessary after a systematic review has reported a clinically important finding but publication bias cannot be safely excluded as an alternative explanation.

Confirmatory large trials may remain important even when prospective registries of trials are available. This is because publication bias is only one of many potential threats to trial validity. When compared with a universal register of all trials, confirmatory large trials are more selective about research areas and objectives, but more flexible in minimising the impact of other biases, for example, biases related to study design and the selection of control interventions, participants and setting.

Summary

Ideally, methods should be designed according to the sources of publication bias. However, as was discussed in chapter 5, publication and related biases are associated with many complicated and inter-related factors, including cultural, economic and psychological. Clearly, the available methods are not sufficient to deal with these factors.

There is little hope that conventional paper journals alone can solve the problem of the selective publication of studies with striking results because of their space limitation and their need to maintain newsworthiness. Electronic publishing has many advantages, such as unlimited space, linkage between references, timely publication, and being relatively cheap to disseminate and preserve. For the purpose of reducing publication bias, peer reviewed electronic journals without limitations on space are required. More importantly, editorial policy needs to be changed to allow the acceptance for publication of clinical trials that are based only on methodological criteria and not on the impact of their findings.

Clearly, the ideal solution to publication bias is the prospective registration of all studies at their inception. Although the universal registration of all studies cannot be realised in the near future, there are many encouraging signs that there will be more such registries established as a result of initiatives from the Government or industry in the UK, and in other countries. Large-scale confirmatory studies may be an alternative that will prevent the consequences of publication bias.

Chapter 7

Reducing or detecting publication and related biases in systematic reviews

Methods that could be useful to reduce, detect or adjust for publication bias in systematic reviews are discussed in this chapter. Literature searching, locating unpublished trials and assessment of the risk of publication bias will be discussed first. Then methods designed to be used in meta-analysis are presented, including those for estimating the number of unpublished studies, the funnel plot and related statistical methods, the method of trim and fill, and more sophisticated modelling methods (a detailed discussion of the modelling methods is available in appendix 6). Finally, the importance of updating systematic reviews is discussed.

Literature searching

Studies with negative or less positive results may be unpublished or published in low circulation journals that are often not indexed in commonly used electronic databases such as MEDLINE. When undertaking a systematic review, a thorough literature search may reduce the possibility of missing studies that are published in low circulation journals or grey literature. A comprehensive study of the state of the art of searching for relevant studies is beyond the scope of this review; only some of the main issues are discussed below.

Several studies have evaluated MEDLINE searches designed to identify randomised controlled trials.[117–124] Sensitivity (recall rate or efficiency rate) is defined as the proportion of the number of trials identified from the total number of relevant trials. Precision is defined as the proportion of the number of trials identified from the total number of references (relevant or irrelevant) retrieved. It is understandable that the sensitivity is inversely correlated with the precision in a MEDLINE search. The reported sensitivity of MEDLINE searches ranged from 16.8% to 97.7% and the precision ranged from 1.6% to 58.8%, depending on the field, search strategy and gold standard adopted.

Adams and co-workers[124] found that 51 of 67 randomised studies not included in MEDLINE were conference abstracts, letters or brief reports. Kleijnen and Knipschild[121] noted that the total number of articles found with computer searches depended strongly on the subject: the MEDLINE database included only 23 of the 107 trials of homeopathy, 28 of the 61 trials of ascorbic acid and the common cold, and 18 of the 45 trials of ginkgo biloba for intermittent claudication and cerebral insufficiency. Odaka and colleagues[273] searched MEDLINE and EMBASE using the same search terms in the domain of Japanese life science. They found variations in the results due to different journals' indexing and recording methods.

The CCTR, a database of controlled trials in health care, is developed, updated and maintained by the Cochrane Collaboration. By the end of 1999, CCTR had included 250,789 references of controlled trials.[231] A study found that the CCTR is very specific but not sensitive for the retrieval of trials in musculoskeletal disorders, because some relevant trials had not been indexed.[274]

A systematic review of published literature relating to "near patient testing" included 102 relevant articles: 49% of these were identified by searching electronic databases (such as BIDS, MEDLINE, EMBASE, CINAHL, PsycLIT), 39% by people working in the field, and 30% by handsearches of abstracts, reports or references.[275]

It is advisable to search more than one electronic database for relevant studies because many journals are indexed in only one of the commonly used databases.[126] In addition, searching electronic databases alone is seldom sufficient; this should be supplemented by checking the references of relevant studies and contacting experts or organisations.[121,275,276] It has been suggested that "a review with a comprehensive search uses at least three sources and provides a description of efforts to identify unpublished trials."[277]

Locating unpublished trials

One study showed that about 31% of published meta-analyses included unpublished data.[278]

The unpublished trials were often identified by surveying individuals, organisations or pharmaceutical companies (*Table 4*).[32,45–50] The number of questionnaires needed for each unpublished study ranged from 1 to 5 in surveys without restrictions on the study area. However, if the purpose was to obtain unpublished studies for a meta-analysis, the number of questionnaires needed for one unpublished study was 173 in Shadish and colleagues' study.[49] In many meta-analyses there may be no unpublished trials identified by surveying potential authors, research funding agents and industry. For example, in a systematic review of near patient testing, no unpublished data were obtained by sending questionnaires to 194 academics and 152 commercial companies.[275]

It is still controversial and unclear how far reviewers should go in identifying unpublished trials. The quality of these trials may be different from those that are published.[267] It was suggested that:

> "authors do not submit their work for publication because they have become aware of serious limitations in it [and, therefore] a meta-analysis is more likely to be accurate if it is based on published studies than if it nets all studies, published and unpublished."[167]

A study of "data on file" cited in pharmaceutical advertisements found that quality assessment of this data was often difficult because much of the material was incomplete.[279]

The use of unpublished data may not necessarily reduce the bias in meta-analysis, particularly if the unpublished data are provided by interested sources such as pharmaceutical companies.[280] Unless one can decide whether the identified unpublished trials represent all the unpublished trials and estimate the proportion of identified unpublished trials in all unpublished trials, the potential for publication bias cannot be satisfactorily solved by locating these trials.

Assessing the possibility of publication bias

Some study characteristics were found to be related to the risk of publication bias, such as observational studies, small sample size and small effect size (see chapter 5). In addition, a comprehensive assessment of study quality is important to detect other potential biases (selection bias, performance bias, attrition bias; and detection bias).[281] Another very important aspect is the conflict of interest of research and funding sources, particularly for deciding the possible direction of bias due to the selective publication of study results. The risk of bias may be great if all trials are funded by a single body having explicit or implicit reasons for favouring a particular finding. Conversely, when similar results are obtained from trials funded by different bodies with conflicting interests, the bias due to funding bodies may become less important.

Funnel plot

Because of a larger random error, the results from smaller studies will be more widely spread around the average effect. If there is no publication bias, a plot of sample size versus treatment effect from individual studies in a meta-analysis should be shaped like a funnel.[2] If the chance of publication is greater for trials with statistically significant results, the shape of the funnel plot may become skewed.

In a funnel plot, the treatment effects from individual studies are often plotted against their standard errors (SEs) (or the inverse of the SEs) instead of the corresponding sample sizes (*Figures 7* and *8*). The use of SEs may have some advantages because the statistical significance is determined not only by the sample size but also by the level of variation in the outcome measured, or the number of events in the case of categorical data. However, the visual impression of a funnel plot may change by plotting treatment effects against SEs instead of against the inverse of the SEs.[282]

Light and Pillemer described two ways in which the shape of the funnel plot can be modified when studies with statistically significant results are more likely to be published.[2] First, assume that the true treatment effect is zero. Then the results of small studies can be statistically significant only when they are far away from zero, either positive or negative. If studies with significant results are published and studies with results around zero are not published, the funnel plot may not be obviously skewed but there will be an empty area around zero (e.g. *Figure 7b*). These polarised results (significant negative or positive results) may cause much debate; however, the overall estimate obtained by combining all studies is unlikely to be biased.

Secondly, when the true treatment effect is small or moderate but not zero, small studies reporting a small effect size will not be statistically

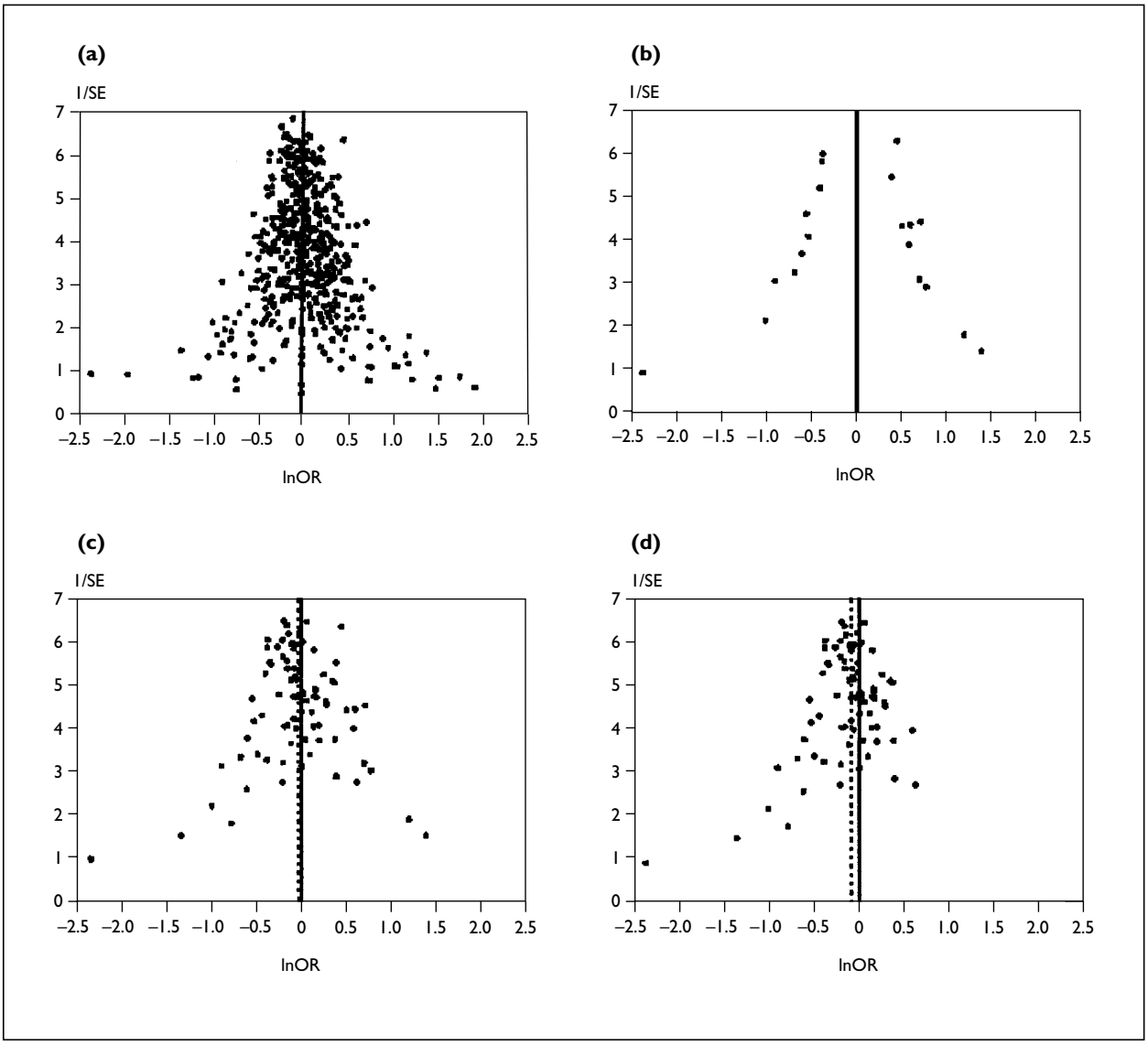


FIGURE 7 *Funnel plot: computer simulations under different assumptions about biased selection when there is no treatment effect*

(a) Without selection bias; trials conducted *n* = 500; trials published *n* = 500; true lnOR = 0.0; estimated lnOR = – 0.007
(b) Selection bias: |Z| > 1.96; trials conducted *n* = 500; trials published *n* = 20; true lnOR = 0.0; estimated lnOR = 0.0114
(c) Selection bias: sample size + (|Z| > 1.96); trials conducted *n* = 500; trials published *n* = 85; true lnOR = 0.0; estimated lnOR = – 0.032
(d) Selection bias: sample size + (Z < – 1.96); trials conducted *n* = 500; trials published *n* = 76; true lnOR = 0.0; estimated lnOR = – 0.085

significant and are therefore less likely to be published, while small studies reporting a large effect size may be statistically significant and more likely to be published. Consequently, there will be a lack of small studies with small effect in the funnel plot and the funnel plot will be skewed with a larger effect among smaller studies and a smaller effect among larger studies (e.g. *Figure 8b*). This will result in an overestimation of the treatment effect in a meta-analysis.

The selection of a study for publication may be a function of many variables, such as sample size, level of statistical significance, extent or direction of difference between comparison groups, and design quality. If the publication of a study is associated with the direction of the results, the extent of publication bias may be much greater than that in which the publication is associated with only the level of statistical significance. For example, if only results with a lnOR of less than zero are selected in *Figure 7b*, the estimated lnOR will be – 0.50 (corresponding to an OR of 0.60), although the true lnOR is zero. *Figures 7c* and *7d* are the funnel plots in which the selection is a function of statistical significance and the

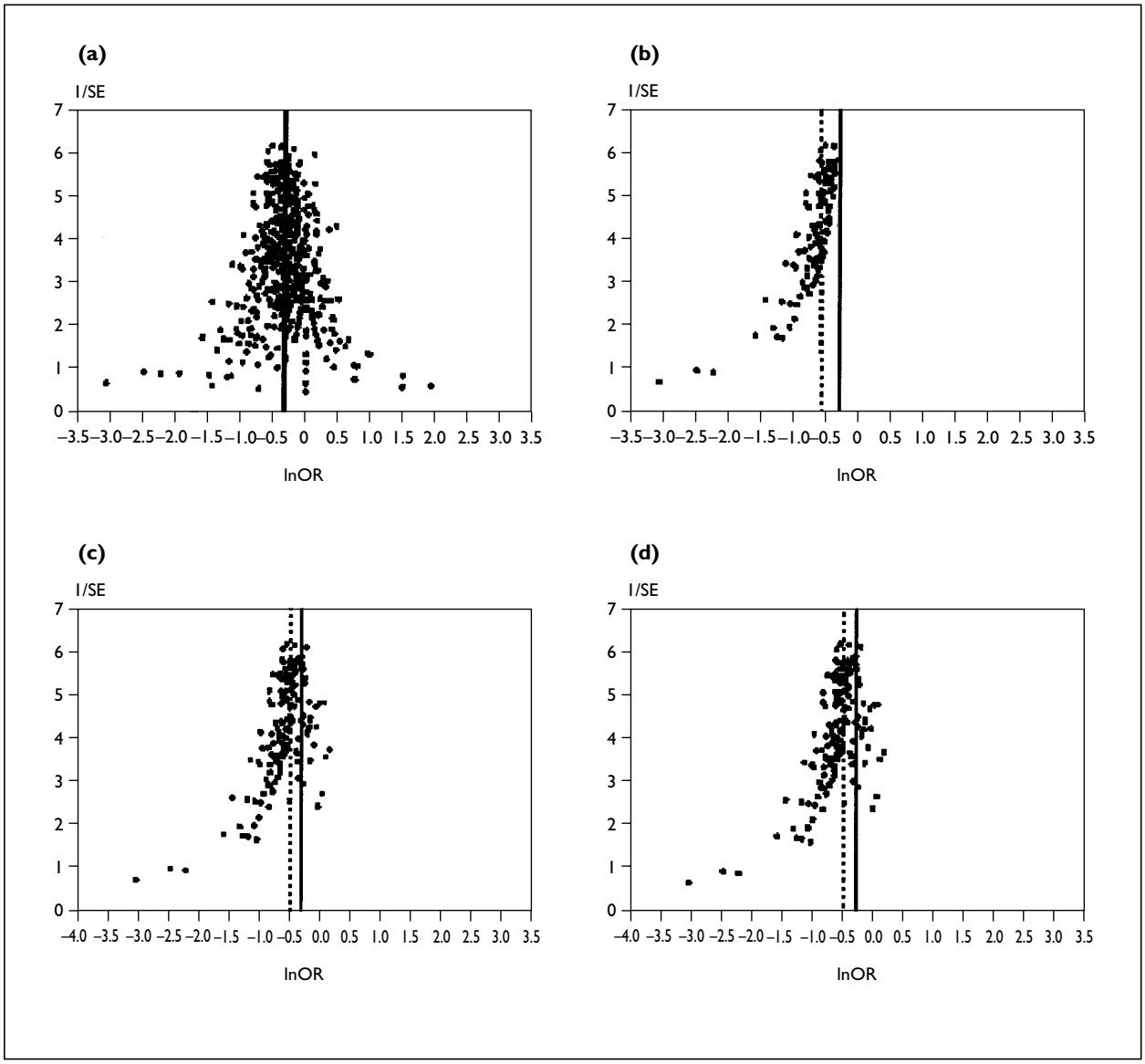


FIGURE 8 *Funnel plot: computer simulations under different assumptions about biased selection when there is a small treatment effect*

(a) Without selection bias; trials conducted *n* = 500; trials published *n* = 500; true lnOR = – 0.315; estimated lnOR = – 0.326

(b) Selection bias: |Z| > 1.96; trials conducted *n* = 500; trials published *n* = 114; true lnOR = – 0.315; estimated lnOR = – 0.588

(c) Selection bias: sample size + (|Z| > 1.96); trials conducted *n* = 500; trials published *n* = 155; true lnOR = – 0.315; estimated lnOR = – 0.49

(d) Selection bias: sample size + (Z < – 1.96); trials conducted *n* = 500; trials published *n* = 155; true lnOR = – 0.315; estimated lnOR = – 0.49

sample size when the true treatment effect is zero. *Figures 8a* to *8d* show the funnel plots of the results of computer simulation under different selection assumptions when there is a small treatment effect (lnOR – 0.315).

There are some limitations in the use of the funnel plot to detect publication bias. For a funnel plot to be useful, there needs to be a range of studies with varying sizes. The funnel plot is an informal subjective method for assessing potential publication bias; different people may interpret the same plot differently. It should also be stressed that a skewed funnel plot may be caused by factors other than publication bias. Other possible sources of asymmetry include different intensity of intervention, differences in underlying risk, poor methodological design of small studies, inadequate analysis, fraud, choice of effect measure, and chance.[44]

Statistical and modelling methods

There are several statistical and modelling methods that can be used to test publication bias in meta-analysis (*Table 8*). Most of the methods could be used for both binary outcomes and continuous outcomes. The discussion in this section will be brief and non-technical. More details about statistical and modelling methods are provided in appendix 6.

Rosenthal's fail-safe N method

Several statistical methods have been developed to estimate the number of possible unpublished studies in a meta-analysis.[5,283–285] The most commonly used method is Rosenthal's fail-safe N or file-drawer method.[5] It is a statistical method to estimate the number of unpublished studies required, with zero treatment effect on average, to overturn a significant result in a meta-analysis.[5] If the number of unpublished studies required to overturn the statistically significant result is large and therefore unlikely to exist, the impact of publication bias is considered to be ignorable and thus the results obtained from published studies to be reliable.

The plausible number of unpublished studies may be hundreds in some areas or only a few in others. Therefore, the estimated fail-safe N should be considered in proportion to the number of published studies (K). Rosenthal suggested that the fail-safe N may be considered as being unlikely if it is greater than a tolerance level of $5K + 10$.[5]

A meta-analysis of ten RCTs of risperidone versus typical neuroleptics in the treatment of schizophrenia shows that risperidone is associated with statistically significantly more patients who had clinically improved (OR 0.75; 95 % CI 0.61 to 0.92) (*Figure 9*).[286] By applying Rosenthal's method, 13 unpublished studies with zero treatment effect on average are required in this meta-analysis to change the statistically significant result into a statistically non-significant result. Although the fail-safe N is greater than the number of published studies, it is much less than 60 (that is, $(5 \times 10) + 10$), a tolerance level suggested by Rosenthal.

Two problems with the fail-safe N method have been identified.[287] First, the method over-emphasises the importance of statistical significance. Secondly, it may be misleading when the unpublished studies have an average effect that is in the opposite direction to the observed meta-analysis. If the unpublished studies reported contrary results compared with those in the published studies, the number of unpublished studies required to overturn a significant result would be smaller than that estimated, assuming an average effect of zero in unpublished studies.

Rank correlation test

Begg and Mazumdar suggested that a rank correlation test can be used to examine the association between effect estimates and their variances, as a complementary method to the funnel plot.[288] The rank correlation test is a distribution-free method, which involves no modelling assumptions (see appendix 6 for more details). However it suffers from a lack of power and so the possibility of publication bias cannot be ruled out when the test is non-significant.

According to simulated results, the power of the rank correlation test is related to several factors: the number of component studies in the meta-analysis, the underlying effect size parameter, the range of variances across studies, the strength of the selection function, and the presence of one-sided or two-sided selection pressures.[288] The test is fairly powerful for large meta-analyses with 75 component studies, but has only moderate power for meta-analyses with 25 component studies. In the meta-analysis of risperidone for schizophrenia, which included ten studies,[286] the rank correlation test did not find an association between the estimated treatment effects and their variances (Spearman's rho correlation coefficient 0.018; $p = 0.96$).

Linear regression approach

Allison and co-workers[41] assessed four published meta-analyses of obesity treatment, using regression techniques to assess the association between treatment effect and sample size. A significant association was observed in two meta-analyses that contained a large number of primary studies (68 and 418 respectively). For the other two, the authors suspected that "there might have been insufficient statistical power to detect publication bias using the funnel plot method." [41]

Egger and colleagues[44] suggested a method to test the asymmetry of a funnel plot, based on a regression analysis of Galbraith's radial plot.[289] The standard normal deviate (SND) is defined as the lnOR divided by its SE. The SND is then regressed against the estimate's precision (the inverse of the SE): $\text{SND} = (a + b) \times \text{SE}(lnOR)^{-1}$. This corresponds to a regression analysis of Galbraith's radial plot.[289]

TABLE 8 *Statistical and modelling methods for publication bias in meta-analysis*

Methods	Basic assumptions	Comments
Rosenthal's fail-safe N	The unpublished studies, on average, have an effect size of zero (z = 0).	If a meta-analysis produces a significant result, this method is used to calculate a "fail-safe N", defined as the number of unpublished studies (defined under basic assumptions) required to overturn the significant result. If the fail-safe N is large and therefore unlikely to exist, the impact of publication bias should not be serious enough to change the conclusion of the meta-analysis (although it may well change the effect size estimate). This is a simple method and easy to use. However, it is not clear how large a fail-safe N should be when the potential publication bias could be ignored. The interpretation of fail-safe N may be difficult. It may be misleading when the unpublished studies have an average effect that is in the opposite direction to the observed meta-analysis rather than a zero effect, which this method assumes. The method combines z-scores and therefore does not directly account for the sample sizes of the studies; hence it is not influenced by the shape of the funnel graph. Heterogeneity between studies is ignored.
Funnel plot	Because of larger random error, the results from smaller studies will be more widely spread around the average effect. A plot of sample size (or variance, or SE) versus treatment effect from individual studies in a meta-analysis should be shaped like a funnel if there is no bias. If the chance of publication is greater for trials with statistically significant results, the shape of the funnel plot may become skewed.	The method is simple. For a funnel plot to be useful, there needs to be a range of studies with varying sizes. The visual impression of a funnel plot may change depending on which measure of trial magnitude (SE, variance, or sample size etc.) is used, or which outcome scale is used (e.g. a funnel plot using the risk difference scale may give a different impression to if the OR scale was used). Furthermore, it is an informal subjective method and the same plot may be interpreted differently by different people. More importantly, an asymmetrical funnel plot may be due to factors other than publication bias, and a symmetrical funnel plot cannot exclude the existence of publication bias.
Rank correlation test	Same premise as the funnel plot. It is a non-parametric test, estimating the correlation between effect size estimates and their variances. No modelling assumptions are required.	This is a method complementary to the funnel plot. It tests the association between effect estimates and their variances in a meta-analysis. It suffers from a lack of power and so the possibility of publication bias cannot be ruled out when the test is non-significant. It is difficult or impossible to determine whether the observed association between the effect estimates and their variances is due to publication bias or other factors also related to study size and outcome.
Linear regression approach	Same premise as the funnel plot. It tests for a linear relationship between each study's SND and the estimate's precision. It assumes: (1) the SNDs are normally distributed for each precision value; (2) the variability in the SNDs is the same for each precision value; and (3) the relationship between SND and precision is linear.	This is a statistical method, based on a regression analysis of Galbraith's radial plot, to test whether a funnel plot is asymmetrical. The operating characteristics of this method need to be evaluated further. When testing the asymmetry of a funnel plot, the gradient of the slope cannot be interpreted as in the original Galbraith radial plot. This linear regression method appears to be more sensitive than the rank correlation method for testing the asymmetry of a funnel plot. Many limitations of a funnel plot also limit the usefulness and interpretation of this method.
Trim and fill method	Same premise as the funnel plot. Assumes the funnel is truncated with studies with the least beneficial estimates, which have been suppressed (the size of studies is not taken into account). The degree of truncation is ascertained by an iterative non-parametric rank-based estimation procedure. A fixed or random effects meta-analysis model can be used to combine studies.	This is a rank-based data augmentation technique utilised to formalise the use of the funnel plot, which is relatively easy to compute. It is able to provide an estimate of the treatment effect by adjusting for potential publication bias in a meta-analysis by imputing suspected missing studies. However, like other methods, the adjusted result will be misleading if the observed asymmetry in a funnel plot is due to factors other than publication bias. The adjusted result is not intended to give a "better" estimate *per se*, but can be used as a form of sensitivity analysis to help to ascertain the likely impact of publication bias on the meta-analysis.

continued

TABLE 8 contd *Statistical and modelling methods for publication bias in meta-analysis*

Methods	Basic assumptions	Comments
Selection models using weighted distribution theory	It is usually assumed that the probability for a study being published is determined by its *p*-value (although sample size could be used instead, but not both simultaneously). The probability that a study is published given that its *p*-value can be defined by the researcher over the range of possible *p*-values, or left to be determined by the model. A fixed or random effect meta-analysis model can be used to combine studies. Study level covariates can be included in the model.	Selection models can be used to adjust a meta-analysis for suspected selected publication. Many different weight functions have been suggested to define the probabilities that studies are published. These methods are still in the experimental stage of development and, like trim and fill, can be used to assess the robustness of meta-analyses results to publication bias. Selection models are quite sophisticated and there is currently a lack of software to implement them. The weight functions suggested are often based on strict assumptions or very limited empirical data, which may be overlooked because of the complexity of these methods. Such methods have been used very rarely in practice, which can be attributed in part to their complexity.
Copas' sensitivity approach	The process of study selection is assumed to be described by a separate regression model with residuals that are correlated with study outcome (assumed to be distributed bivariate normally). The probability that a study is published relates to a regression equation, with a gradient determined by a factor of the square root of the study sample size. Both outcome and sample size are included in the modelling equations. A random effect meta-analysis model is used. Study level covariates can be included in the model.	This method provides a sensitivity analysis for examining the potential extent of publication bias. The equation that the method derives for the pooled estimate contains too many unknown terms that cannot be estimated without large assumptions. This is why a sensitivity analysis approach is advocated, which allows an assessment of different levels of publication bias on the pooled estimate. This method has only recently been described and, owing to its complexity, software is required before it can be used routinely by those carrying out systematic reviews.

If there is no selection bias, the points from individual trials will scatter about a line that runs through the origin at standard normal deviate zero ($a = 0$), with the slope (b) indicating the size and direction of effect. If there is asymmetry, with smaller studies showing effects that differ systematically from larger studies, the regression line will not run through the origin. The intercept (a) provides a measure of asymmetry; the larger its deviation from zero, the more pronounced the asymmetry. Negative values ($-a$) will indicate that smaller studies show big protective effects. It was recommended that both weighted and unweighted regression analysis should be carried out, and that the result that showed the greatest deviation from zero should be selected. Considering the relative low power for all heterogeneity tests, a 10% level of significance was recommended.[44]

By applying this method, significant asymmetry was observed in 38% of 37 meta-analyses published in a selection of journals and in 13% of 38 Cochrane reviews.[44] Egger and colleagues also identified four meta-analyses in which discordant funnel plots showed that the treatment effect was larger in the meta-analyses than in the corresponding large trials. Using these workers' method, significant asymmetry was found in three of these four meta-analyses. When the rank correlation test was used, only one of the four showed significant asymmetry. Thus the linear regression method[44] appears to be more sensitive than the rank correlation test.[288]

Figure 10 shows the results of applying this method to the meta-analysis of risperidone for schizophrenia. Because the intercept of the weighted regression is significantly less than zero, it indicates that the small studies are associated with a larger treatment effect.

In the original Galbraith radial plot, the slope of the line indicates the size and direction of effect; a greater gradient of the slope indicates a greater difference in the treatment effect.[289] However, when testing the asymmetry of a funnel plot, the gradient of the slope will become closer to zero or positive when the

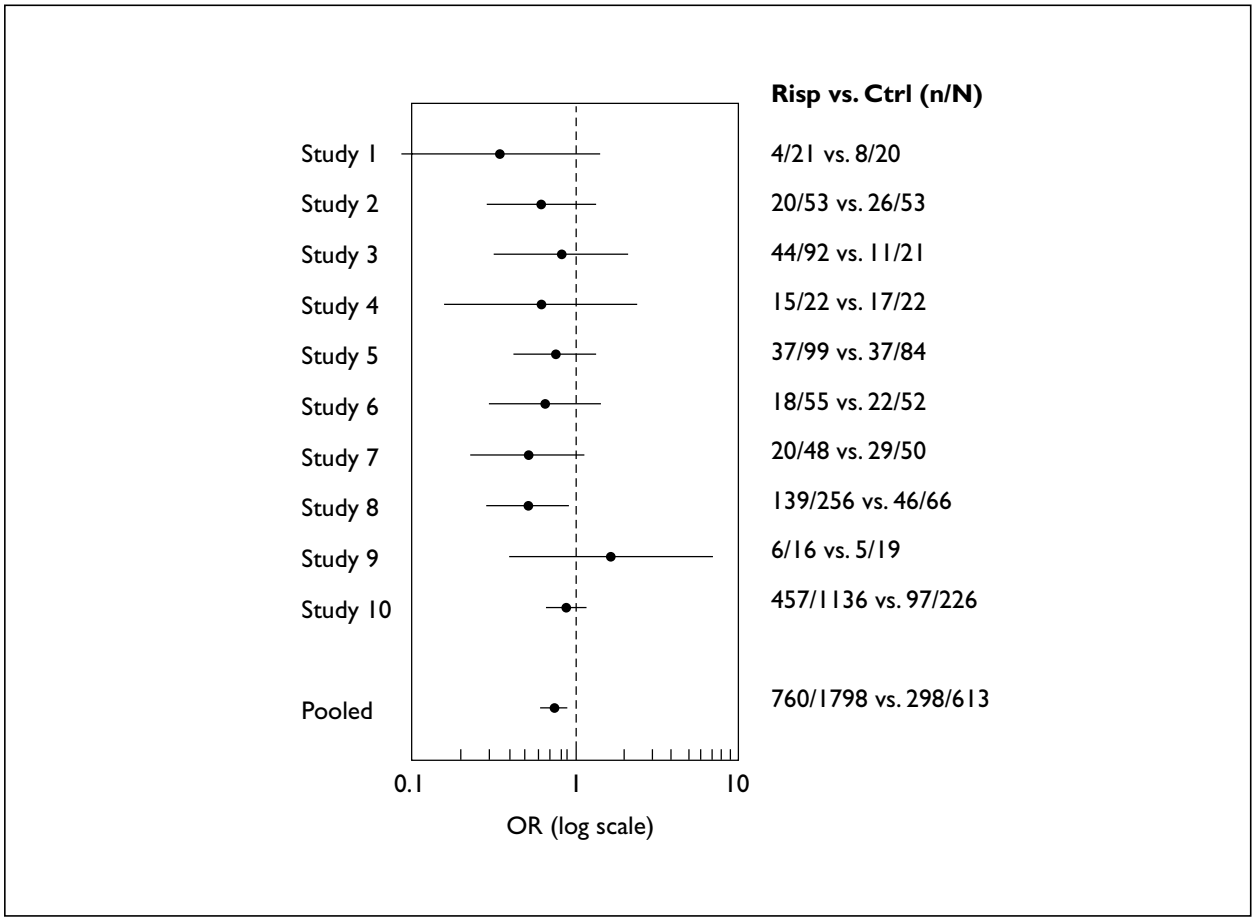


FIGURE 9 *Meta-analysis of risperidone for schizophrenia: number of patients classified as clinically not improved (n) with risperidone (Risp) versus typical neuroleptic medication (Ctrl); data from Kennedy* et al. [286] *(N, number of patients in treatment group)*

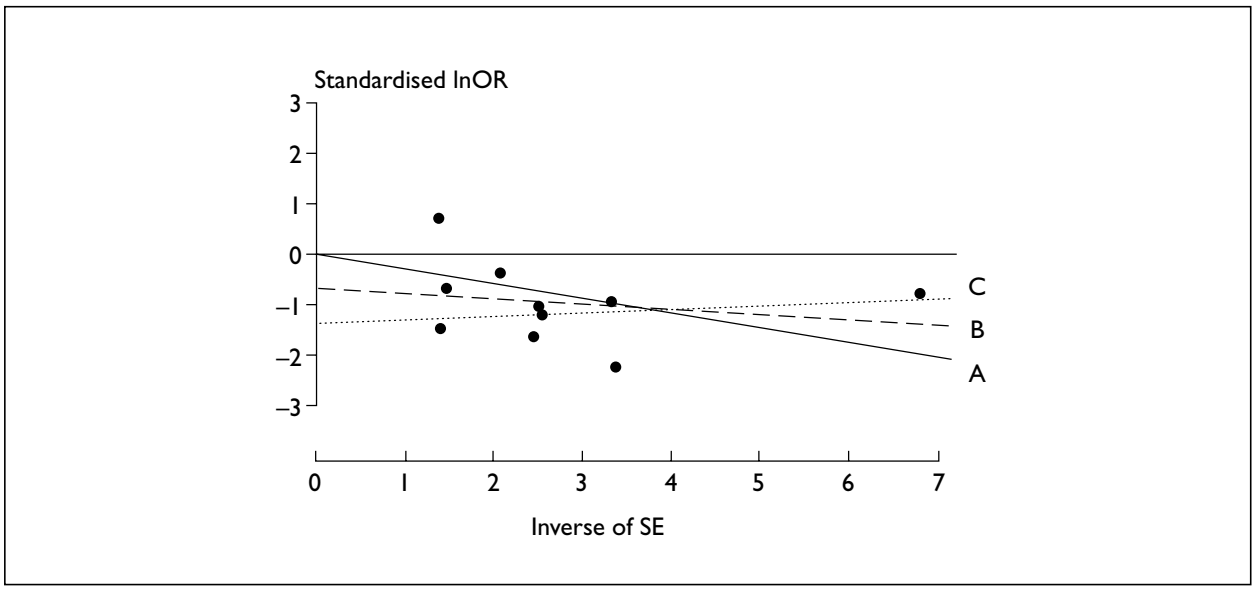


FIGURE 10 *Linear regression method to test asymmetry of funnel plot: meta-analysis of risperidone for schizophrenia as an example. The slope of line A corresponds with the pooled lnOR if there is no bias (i.e. the intercept = 0). Line B is an unweighted regression in which the intercept is negative but not statistically significant (intercept = – 0.704,* p = *0.223). Line C is a weighted (by the inverse of variance) regression in which the intercept is significantly less than zero (intercept = – 1.392,* p = *0.016), therefore it indicates the existence of bias in favour of small trials.*

estimated effect is greater in smaller studies. The operating characteristics of this method need to be evaluated thoroughly by more analytical work or by computer simulations.[238]

Trim and fill method

The trim and fill method is a simple rank-based data augmentation technique to formalise the use of the funnel plot.[282] This recently developed method can be used to estimate the number of missing studies and, more importantly, to provide an estimate of the treatment effect by adjusting for potential publication bias in a meta-analysis. Briefly, the asymmetrical outlying part of the funnel is first "trimmed off" after estimating how many studies are in this part. Then the symmetrical remainder is used to estimate the "true" centre of the funnel. Finally, the "true" mean and its variance are estimated based on the "filled" funnel plot in which the trimmed studies and their missing "counterparts" symmetrical about the centre are replaced. In simulation studies, it was found that this method estimated the point estimate of the overall effect size approximately correctly and the coverage of the CI is substantially improved compared with ignoring publication bias.[282]

Applying the trim and fill method to the meta-analysis of risperidone for schizophrenia (*Figure 11*) suggests that this funnel plot would need three trials to fill the right side. After filling with the assumed missing trials, the adjusted OR (0.82; 95% CI 0.68 to 1.01) becomes non-significant compared with that based on the published trials (0.75; 95% CI 0.61 to 0.92). The advantage of risperidone versus conventional neuroleptics in clinical improvement for schizophrenia becomes smaller and no longer statis-tically significant after adjusting for potential publication bias.

Sophisticated modelling methods

The impact of missing studies may also be assessed by using more experimental and sophisticated modelling methods. Because of their nature, these methods are discussed only briefly below; more detailed treatment is available in appendix 6.

Several selection modelling methods have been developed to investigate or adjust the results of a meta-analysis in the presence of publication bias. Many of these methods are related and based on weighted distribution theory derived from both classic[204,290–293] and

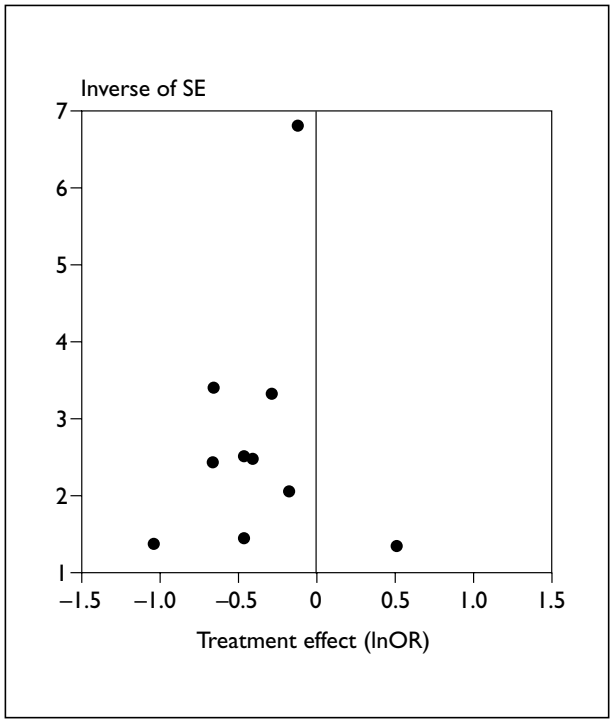


FIGURE 11 *Funnel plot of lnOR against inverse of SE: meta-analysis of risperidone versus typical neuroleptics for schizophrenia*

Bayesian[294,295] perspectives, although other methods do exist.[296]

There are two aspects to selection models that use weighted distribution theory: an effect size model that specifies what the distribution of the effect size estimate would be if there were no selection, and the selection model that specifies how this effect size distribution is modified by the selection process.[297] In some methods the nature of the selection process is predefined by the researcher, while in others it is dictated by the available data.

Unfortunately, the complexity of these methods means that they have largely been used only by statistical modelling experts. Hence, although applications of their use do exist,[295,298] they are limited in number and no comprehensive comparison of the methods has ever been carried out. For this reason, many feel they are still in the experimental stage of development, despite being considered for well over a decade.[205] The development of user-friendly software is required to bring these methods into more mainstream use. They have the potential to be a valuable tool for assessing the robustness of meta-analyses results to publication bias, if not to adjust the pooled estimate *per se*.[205]

Comparison of different statistical and modelling methods

Pham and co-workers identified 31 methods, including seven file-drawer methods, nine funnel plot related methods, 11 methods based on a selection model, and four methods of a selection model with data augmentation.[299] These methods were compared using 26 meta-analyses that included 327 published RCTs and 73 unpublished RCTs. It was found that treatment efficacy was overestimated on average by 6% (interquartile range – 3% to 43%) by published trials. On average the available methods tend to overcompensate the publication bias and underestimate the treatment effect. The treatment effect was underestimated by 6% (interquartile range – 39% to 18%) by the trim and fill method, 21% (– 19% to 76%) by Egger's linear regression method, and 47% (– 22% to 104%) by a selection model. They concluded that the methods for dealing with publication bias may provide different estimates in terms of the direction and extent of publication bias.[299] This is the first study that aimed to compare different methods by using an "unbiased sampling frame". However, the results of this study should be interpreted with caution because the inclusion of some unpublished trials in the meta-analyses may not necessarily mean that the sample of trials is unbiased.

Fixed or random effects models

In meta-analyses, larger studies will give greater weight than smaller studies when results are quantitatively combined. This procedure will have an advantage in reducing the impact of publication bias because less weight is given to smaller studies that are associated with a greater risk of publication bias.

There are two statistical models that can be used to combine the results of individual studies in a meta-analysis: the fixed effects model and the random effects model.[300] In the fixed effects model it is assumed that all individual studies are measuring a single value of the true effect and that the observed difference between the studies is due to sampling error. The precision (e.g. the inverse of within-study variance) of individual results is employed as the weight for each study to estimate the overall result in meta-analysis using the fixed effects model. On the other hand, the random effects model assumes that individual studies are measuring a range of possible effects. In addition to the variation within studies, the variation between studies is also incorporated into a meta-analysis by using the random effects model. In reality, the differences between the fixed effects model and the random effects model are often ignorable.[301] When there is considerable heterogeneity between individual studies the CI estimated by using the random effects model will be wider than that by using the fixed effects model.

The weights used to combine individual studies are the inverses of within-study variances in the fixed effects model, and are the inverses of within- plus between-study variances in the random effects model. Therefore, by giving relatively larger weights to smaller studies, the random effects model is more vulnerable to publication bias than the fixed effects model.[302]

Updating systematic reviews

Considering the likely delayed publication of studies with negative or less favourable results, it is important to update systematic reviews when new evidence becomes available. Jadad and colleagues[303] compared 36 Cochrane reviews with 39 meta-analyses published in paper-based journals (randomly selected sample). They found that more Cochrane reviews included a description of the inclusion/exclusion criteria (35/36 versus 18/39; $p < 0.001$) and assessed trial quality (36/36 versus 12/39; $p < 0.001$). No Cochrane reviews had language restrictions (0/36 versus 7/39; $p < 0.01$). Within 2 years of publication, 18 of the 36 Cochrane reviews had been updated versus only one of the 39 reviews published in paper-based journals. It was concluded that "Cochrane reviews appear to have greater methodological rigour and are more frequently updated than systematic reviews or meta-analyses published in paper-based journals."[303]

Possible reasons given for a very low update rate among paper-based reviews included editors of such journals not being sufficiently interested in publishing updated versions of previously published systematic reviews, authors not being aware of such interest, or authors lacking the interest or resources to update previously published reviews.[303]

Summary

Many methods have been suggested for preventing or testing for publication bias.

The available methods could be classified as those for preventing publication bias (discussed in chapter 6) and for dealing with publication bias in systematic reviews (discussed in this chapter). In addition, it is possible to classify the available methods according to the stage of a literature review: to prevent publication bias before a literature review (e.g. prospective registration of trials), to detect publication bias during a literature review (e.g. locating unpublished studies, fail-safe N, funnel plot, modelling), or to minimise the impact of publication bias after a literature review (e.g. confirmatory large-scale trials, updating the systematic review).

This chapter has discussed the methods available for dealing with publication and related biases in systematic reviews, including literature searching, locating unpublished studies, assessment of the risk of publication and related biases, several methods for testing or adjusting for publication bias in meta-analysis, and updating systematic reviews. There may be different opinions about the importance of development and the use of analytical methods for detecting and/or adjusting for publication bias in meta-analysis. It should be stressed that these methods are by nature indirect and exploratory, and often based on certain strict assumptions that can be difficult to verify in the real world (*Table 8*). The attempt at identifying or adjusting for publication bias in a systematic review should mainly be used for the purpose of sensitivity analyses.

Chapter 8

Survey of published systematic reviews

Studies on publication bias may themselves be as vulnerable as clinical trials to publication bias in favour of positive findings.[7] It is hoped that published systematic reviews in DARE at the NHS Centre for Reviews and Dissemination, the University of York, may provide an unbiased sample to examine publication bias because reviews are included in DARE according to broad quality criteria.

At the end of August 1998, 193 systematic reviews published in 1996 were included in DARE. These reviews have been selected to examine the issues and methods relevant to publication bias. *Table 9* presents the main results of the survey.

Among the 193 systematic reviews, the majority (83%) evaluated the effectiveness of health interventions; 8% focused on the adverse effects of interventions; and 9% evaluated diagnostic technologies. One hundred and thirty-two of these reviews were meta-analyses, in which the results of primary studies were combined quantitatively, while 61 were narrative systematic reviews, in which quantitative pooling was not employed.

Literature searching in the reviews

The databases used for identifying relevant studies were not reported in five of the 193 included reviews. MEDLINE and bibliographies were most commonly searched (*Figure 12*). Fourteen reviews used a MEDLINE search alone and one EMBASE alone. For the majority of the reviews, two or more sources were searched; 69% (n = 133) of the reviews resulted from a combined a search of MEDLINE with the checking of bibliographies to identify relevant studies. Other methods used for literature searching included: contacting experts or authors (26%) or companies (9%), searching conference abstracts (18%), EMBASE (17%), PsycLIT (11%), CINAHL (8%), BIDS (7%), Current Contents (5%), Cochrane databases (5%), and handsearching selected journals (15%). In addition to these general purpose databases, a review on a specific topic may search the specialised databases. For example, for three reviews of cancer treatments CANCERLIT was searched. The authors of a review of low molecular weight heparins searched a prospective registry of trials

TABLE 9 *Publication bias: a survey of 193 DARE systematic reviews*

	No. narrative reviews (%)	No. meta-analyses (%)	Total no. (%)
No. reviews	61	132	193
Unpublished studies			
Explicitly sought	9 (15)	43 (33)	52 (27)
Searched for or included	18 (30)	47 (36)	65 (34)
Non-English literature			
Explicitly sought	11 (18)	25 (19)	36 (19)
Searched for or included	14 (23)	43 (33)	57 (30)
Discussion of publication bias in the review	6 (10)	63 (48)	69 (36)
Use of methods for publication bias	0 (0)	33 (25)	33 (17)
Reviews' conclusions			
Significant/positive	25 (41)	90 (68)	115 (60)
Non-significant/negative	8 (13)	18 (14)	26 (13)
Unclear	28 (46)	24 (18)	52 (27)

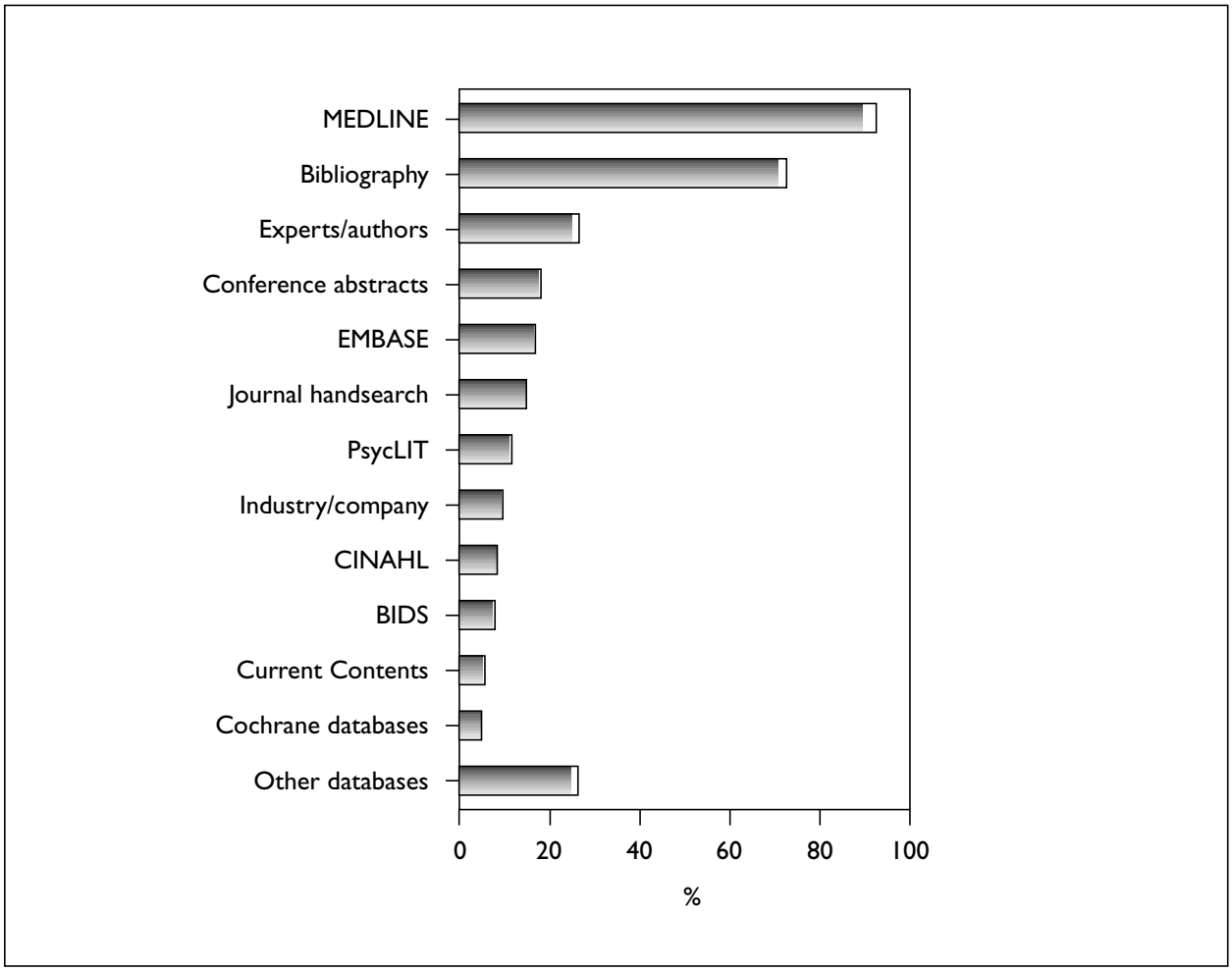


FIGURE 12 *Databases searched for literature in 193 DARE reviews*

(Registries of the International Society of Thrombosis and Haemostasis).[304] The database of Physician's Data Query was used for three reviews by the Swedish Council on Technology Assessment in Health Care.[305–307]

Unpublished studies

Searching for unpublished studies was reported more often in meta-analyses than in narrative systematic reviews (33% versus 15%). The methods used were mainly writing to investigators/authors, research organisations or pharmaceutical companies. Meeting proceedings were often searched to identify unpublished abstracts if conference abstracts were considered as unpublished. By checking the reference lists of all 193 reviews, it was found that conference abstracts were included in some reviews for which unpublished studies were not explicitly searched. In total, 36% of meta-analyses and 30% of narrative systematic reviews explicitly searched for, or included, studies that were unpublished or only presented as abstracts (*Table 9*).

Seventeen of the 132 meta-analyses explicitly excluded unpublished studies or abstracts because they were not peer reviewed and were therefore considered to be unreliable.

Non-English language studies

Non-English language literature was explicitly sought in 19% of the systematic reviews, with no difference between meta-analyses and narrative systematic reviews. By checking the reviews' references, it was found that non-English language studies were included in a further 11% of the reviews. In total, about 33% of the meta-analyses and 23% of the narrative systematic reviews searched for or included non-English language

studies. However, language restrictions may still exist in many of the reviews that included non-English language literature. For example, reviews were often limited to publications in the major European languages, or in English plus the reviewers' native languages.

Use of methods for publication bias

The problem of potential publication bias was discussed or mentioned more frequently in the meta-analyses than in the narrative systematic reviews (48% versus 10%). Methods for testing publication bias were used in 33 of the 132 meta-analyses and in none of the 61 narrative systematic reviews (*Table 9*). The most commonly used methods were the fail-safe N (14 reviews) and methods such as funnel plots for testing the association between sample sizes and treatment effects (11 reviews). Five systematic reviews compared the results of full articles and the results of abstracts. Other methods used included sensitivity and subgroup analysis. It should be recognised that some methods (e.g. Egger's linear regression method) became available only recently and could not have been used in the reviews published in 1996.

Evidence provided on publication bias

Evidence about the existence or absence of publication bias was available in 33 systematic reviews that tested publication bias (*Table 10*). In 15 of these reviews, the evidence suggested that the results were unlikely to be affected by publication bias; in 11 it was uncertain, and in seven the evidence suggested possible publication bias.

The funnel plot was asymmetrical in six systematic reviews, indicating the possible existence of publication bias. Correlation analysis in three reviews did not find a significant association between sample size and treatment effect. Although the fail-safe N method was most frequently used (14 reviews), it seems difficult to interpret the fail-safe N estimated. For example, five of the 14 reviews that used the fail-safe N method did not discuss or interpret the fail-safe N estimated (*Table 10*). In other reviews, the fail-safe N estimated was often used to indicate that a significant result was unlikely to be affected by publication bias.

Conclusions of surveyed systematic reviews

In assessing whether a review's conclusion was significant/positive, non-significant/negative, or unclear, there was moderate agreement between reviewers (kappa value 0.47). Significant or positive conclusions were made in 42% of the narrative systematic reviews and 68% of the meta-analyses. The proportion of reviews with non-significant or negative conclusions was similar, 13% and 14% respectively, in the narrative systematic reviews and meta-analyses. An uncertain conclusion was reported more often in the narrative systematic reviews than in the meta-analyses (45% versus 18%). Of reviews with significant or positive conclusions, 48% of the meta-analyses and 12% of the narrative systematic reviews discussed or tested for publication bias.

Summary

The results of this survey are consistent with the results from other similar studies. For example, Cook and colleagues[278] found that 31% of 150 meta-analyses included unpublished studies. Another study showed that 28% of a sample of 132 meta-analyses included grey literature.[308] Brazier assessed 165 systematic reviews published between 1985 and 1997,[309] and found that 41% of the systematic reviews included unpublished studies, 47% discussed publication bias, and 17% explicitly stated that there were no language restrictions. Brazier's survey also showed that the proportion of reviews that include unpublished studies or non-English language literature is increasing.[309]

The results of the survey of DARE suggest that potential publication bias has been largely ignored, and that available methods for dealing with it have not been used in most of the systematic reviews surveyed. The problem of publication and related biases has been dealt with more often in reviews containing a meta-analysis than in narrative systematic reviews, although this may simply be a reflection of marked heterogeneity or other factors that make it impossible to conduct a meta-analysis. In particular, there is a lack of methods for dealing with publication bias in narrative systematic reviews.

Fifteen (45%) of the 33 meta-analyses that tested publication bias concluded that bias was unlikely, while the other 55% concluded that it was possible

TABLE 10 *Methods and results of assessment of publication bias in 33 of 193 systematic reviews surveyed*

DARE Accession no.: intervention	Methods and results/conclusions about publication bias
DARE-960384: The risk of stroke and death due to endarterectomy for symptomatic carotid stenosis	Outcome assessment by a neurologist was associated with a higher risk of stroke and death than that by surgeons. Studies with single-surgeon authors were associated with a lower risk of stroke and death than those in which the authorship comprised two or more surgeons.
DARE-960515: Prophylactic nimodipine for delayed ischaemic deficit after subarachnoid haemorrhage	Funnel plot showed that the meta-analysis may be affected by publication bias. Fail-safe N 31 (meta-analysis included six trials). Authors concluded that publication bias was unlikely to affect the significant effects markedly.
DARE-960755: *Helicobacter pylori* eradication for duodenal and gastric ulcer recurrence	Overall ulcer recurrence rate was higher in abstracts than in published articles. These might be explained by the existence of publication bias or methodological flaws in the abstracts.
DARE-960987: The effect of desogestrel and ethinyloestradiol on plasma lipids in healthy women	A large bias seems unlikely because the results in four large efficacy and safety studies in which lipids were a subanalysis of desogestrel are generally consistent with the overall results of this meta-analysis.
DARE-961047: Emergency medical services for victims of out-of-hospital cardiac arrest	Funnel plot did not demonstrate evidence of publication bias (it was not presented in the review).
DARE-961063: Functional electrostimulation in post-stroke rehabilitation	Fail-safe N 90 (four trials in the meta-analysis). Authors considered publication bias not to be an important factor affecting the conclusions.
DARE-961089: Prevention of myocardial infarction and death by antidyslipidaemic therapy	Assuming that an unpublished hypothetical study showed a 10% treatment-induced increase in mortality, this hypothetical study would need to have 10,220 participants to eliminate the significance in total mortality observed with antidyslipidaemic therapy in this meta-analysis with 25 trials and a study population of 18,452.
DARE-961133: Granulocyte-colony stimulating factor for the prophylaxis of neutropenic fever in patients with small cell lung cancer receiving myelosuppressive antineoplastic chemotherapy	Fail-safe N 21 (three trials in the meta-analysis). No further explanation about the fail-safe N estimated.
DARE-961164: Short-term treatment of gastric ulcer	Fail-safe N 2–155, depending on the intervention and outcomes measured. No further explanation about the fail-safe Ns estimated.
DARE-961423: Radiographic efficacy and clinical tolerance; loversol versus other non-ionic contrast media	Although not statistically significant ($p > 0.6$), results were less favourable in unpublished than in published trials.
DARE-961708: Sulodexide therapy in peripheral occlusive arterial disease	Fail-safe N 14–93, depending on the outcomes measured (19 studies included in the meta-analysis). No further explanation about the fail-safe Ns estimated.
DARE-961724: Interferon in chronic hepatitis C	No significant difference was observed when meta-analysis included or excluded evidence presented in abstracts.
DARE-961784: Fluoroquinolones for bacterial infections in neutropenic patients	Fail-safe N 28–196 in two separate meta-analyses including six and 13 studies respectively. No further explanation about the fail-safe Ns estimated.
DARE-961790: Calcium and colorectal neoplasia	Funnel plot indicated that small studies with RRs close to 1.0 were less likely to be reported.
DARE-961862: Methotrexate dose intensity on high-grade osteogenic osteosarcoma	There was no correlation between the sample size of trials and patient outcome.
DARE-961872: Computer-based clinical reminder systems for preventive care in the ambulatory setting	Funnel plot and correlation analysis. The magnitude of the ORs was not significantly correlated with the SEs of the lnORs ($p = 0.24$).
DARE-961915: Psychoeducational care in adults with chronic obstructive pulmonary disease	Fail-safe N was > 50 for six of the ten outcomes estimated (65 studies included in the meta-analysis). Authors concluded that "fail-safe Ns this large provide reasonably strong evidence against a threat to validity for these 6 outcomes based on publication bias". Fail-safe N was 25–35 for the other four outcomes.

continued

TABLE 10 contd *Methods and results of assessment of publication bias in 33 of 193 systematic reviews surveyed*

DARE Accession no.: intervention	Methods and results/conclusions about publication bias
DARE-963960: Quality of life in cardiac patients	Fail-safe N 16,766 (84 studies in the meta-analysis). No further explanation about the fail-safe N estimated.
DARE-964342: Psychoeducational care in adults with asthma	Fail-safe N was > 50 for all but two outcomes. Author concluded that "fail-safe N this large provide reasonably strong evidence against a threat to validity based on publication bias".
DARE-965235: Occupational therapy for older persons	Fail-safe N 110 (15 studies in the meta-analysis). Authors concluded that publication bias cannot explain the global positive result because of the following reasons: (1) given the current area of investigation, the existence of 110 unpublished studies does not seem possible; (2) non-significant results may be more publishable in this case; and (3) many excluded studies (e.g. owing to a lack of adequate statistical reporting) were associated with a positive effect.
DARE-965403: Patient education intervention in osteoarthritis and rheumatoid arthritis	Funnel plot was symmetrical for patient education trials, which suggested the absence of publication bias, but asymmetrical for the NSAID trials, which suggested that smaller NSAID trials with little or a negative effect were under-represented in the meta-analysis.
DARE-968048: Stress ulcer prophylaxis in critically ill patients	Inclusion or exclusion of non-English language studies may be one of several reasons for the discrepant results between two meta-analyses.
DARE-968203: Polymerase chain reaction for the diagnosis of HIV infection	Studies published only as abstracts provided lower estimates of sensitivity and specificity. This may indicate publication bias.
DARE-968221: Reduced dietary sodium on blood pressure	Funnel plot was asymmetrical, indicating that small trials showing no effect were under-reported.
DARE-968245: Commercial serological kits for Helicobacter pylori infection.	There was no significant difference in estimated accuracy between abstracts and full articles.
DARE-968370: St John's wort for depression	Several trials were published more than once without reference to previous publication. St John's wort is highly popular in German-speaking countries and virtually unknown in the English-speaking world. No study could be identified when only English language literature was searched.
DARE-968398: Adding heparin to aspirin for reducing myocardial infarction and death in unstable angina	Rank correlation: no significant association between study size and treatment effect (p = 0.64).
DARE-968492: Prevention of nonsteroidal anti-inflammatory drug-induced gastrointestinal mucosal injury	Funnel plot suggested a possible publication bias in the case of gastric lesions. Fail-safe N = 199 short- and 14 long-term trials (six and four trials in the meta-analysis respectively). Publication bias could hamper the dimension of the positive effect.
DARE-968497: Sampling device and detection of abnormality in cervical smears	Funnel plot suggested publication bias in favour of the spatula plus cytobrush (plot not presented).
DARE-970032: Effects of thyroid hormones on bone mass	Funnel plot suggested no major publication bias. Fail-safe N 1–82. Authors recognised that a small number of unpublished studies with diverging results would change the findings of many of the 25 meta-analyses.
DARE-970167: Effect of physician profiling on utilisation	Fail-safe N 252 or 85 or – 0.7, respectively, for three levels of meta-analysis (the corresponding number of trials included was 12, eight and five). Authors compared estimated fail-safe N and the threshold suggested by Rosenthal. (Note: a negative fail-safe N makes no sense.)
DARE-970245: Low molecular weight heparins versus unfractionated heparin in the initial treatment of deep venous thrombosis	The results from several ongoing trials were not available, but it is unlikely that their results will dramatically affect the analysis.
DARE-978099: Hypertension in elderly people	Fail-safe N 84 (16 trials in the meta-analysis). Sensitivity analysis by excluding smaller and poor quality studies did not change the results.

or uncertain. These conclusions concerning the existence of publication bias were all based on indirect evidence and therefore may be unreliable. For example, an observed asymmetrical funnel plot or differences between abstracts and full articles may be explained by factors other than publication bias. On the other hand, publication bias cannot be safely ruled out in the meta-analyses that reported a symmetrical funnel plot or a large fail-safe N value.

In summary, this survey indicates that the literature searching was clearly inadequate in some published systematic reviews. Potential publication bias has been ignored and available methods for dealing with this problem have not been used in most of the published systematic reviews. When they are used to estimate possible publication bias at the stage of literature review, the available methods are far from adequate and their usefulness is strictly limited.

Chapter 9

Discussion and recommendations

The empirical evidence demonstrates that studies with significant or favourable results are more likely to be published or cited than those with non-significant or unfavourable results. The extent and direction of such selective publication is still uncertain and may vary greatly depending on circumstances. In addition, there is little empirical evidence about the impact of publication and related biases on health policy, clinical decision making and the outcome of patient management. Because of the lack of empirical evidence about the impact of publication bias, there is disagreement on the actual importance of publication bias in the evaluation of health care interventions. Some argue that positive studies are more important than negative studies and that the selective publication of positive studies is not a clinical problem,[310–312] while others have argued that the potential consequences of publication bias are serious.[7,143,313] Indeed, we have identified a few well-documented examples that demonstrated detrimental consequences of publication bias (see chapter 4).

It seems that the most common reason for publication bias is that investigators fail to write up or submit studies with non-significant results. Investigators may lose interest in non-significant results or be motivated not to publish results that are unfavourable to their own interest. However, it is possible that investigators would not submit studies with non-significant results mainly because of anticipated rejection by journals. Paper-based journals often select manuscripts that are original and have an important impact on practice. The potential role of research sponsors needs to be emphasised because evidence has shown that studies with results that are favourable to the funding body's interest are more likely to be disseminated.

The existence of publication bias is demonstrated by showing that the publication of studies is associated with the strength or direction of the findings. It is therefore important to define what we mean by two basic concepts: publication and study findings. The formats of publication include full publication in journals, presentation at scientific conferences, reports, book chapters, discussion papers, dissertations or theses. The studies presented only at scientific meetings are often not considered to be published and there are disagreements about how reliable they are. For example, 17 of the 132 meta-analyses surveyed in chapter 8 explicitly excluded abstracts because they were not peer reviewed.

Study findings are commonly classified as being statistically significant or statistically non-significant. Sometimes, results are classified as being negative versus positive,[34,35,57,112] supportive versus unsupportive,[133,195] or striking versus unimportant.[54] The classification of study findings is often dependent upon subjective judgement and may therefore be unreliable. For example, negative results may be defined as those that fail to show the benefit of experimental intervention, no matter whether the experimental intervention is equal to or less effective than the control intervention. However, it may be deemed as a null result when there is no statistically significant difference, and as a negative result only when there exists a statistically significant difference in favour of the control.[314]

In addition to the difficulties in defining publication and classifying outcomes, studies of publication bias themselves may be as vulnerable as other studies to the selective publication of significant or striking findings.[7,315] In a letter to the Editor about Easterbrook and co-workers' study of publication bias,[51] Johnston and Breimer[316] questioned whether it would have been accepted for publication if no significant difference had been found. It is therefore sensible to interpret the available evidence about publication bias with caution.

Publication ethics

Publication ethics could be defined as "a set of expectations about the proper method for preparing and presenting data and information within the literature".[317] The issues that have been discussed in the literature about publication ethics include fraudulent results, duplicate publication, plagiarism, honorary authorship, and the misleading use of statistics.[318–320] It has been suggested that the publication of the results of scientific

research should be an ethical imperative in order to achieve the goal of advancement of human well-being.[318] Under-reporting research is scientific misconduct that may cause inappropriate patient care. It is also unethical to mistreat the trust of the patients involved and waste invested resources.[16,143] Because publication bias may be considered as an ethical issue, it has been suggested that RECs should require the registration of clinical trials at their inception as a condition of approval, and audit the reporting of results of research that has been previously approved.[256]

It seems that the potential consequences of publication bias have been overlooked sometimes by leading experts and there may be an erroneous impression that negative or null findings are less important. For example, it has been suggested that non-significant research "clutters up the journals and does more harm than good to the author's reputation in the minds of the discerning."[321] In a chapter about ethics in the dissemination of new knowledge, Meslin incorrectly recommended that "it is preferable to publish positive research findings, because they advance knowledge", although the same author mentioned preferential bias to publish only positive findings.[317]

The importance of "negative findings" has been stressed by some authors. Chalmers suggested that the term "negative trial" should be outlawed because "all trials that have been well conceived and well conducted – whatever their results – represent positive contributions to knowledge."[322] This suggestion is backed up by Gluud, who believes that "negative trials" are positive.[323] Halperin argued that "just as important as positive trials are the negative trials that show paths that ought not be taken."[324]

Publication bias and other sources of bias

The validity of research may be threatened by many different types of biases, such as patient selection bias, patient allocation bias, and outcome measure bias. Begg and Berlin[7] have recognised that the biases observed in the published literature may be generated from a combination of sources, and it is very difficult to distinguish the effect of publication bias from other biases. It is possible that the existence of other biases can worsen publication bias. For example, studies of poor quality may tend to produce wide-ranging results with some extreme or unusual findings because of a greater risk of biases in subject selection, allocation and outcome measurement. Such unusual findings may be more likely to be published.

The phenomenon that small studies are more vulnerable to publication bias is due to greater random error in the results of small studies. It is also possible that small studies tend to be poorly designed and associated with additional biases. Using a meta-analysis of epidemiological studies as an example, Petticrew and colleagues[325] found that effect size was associated with both the sample size and the study quality. After adjusting for the quality of an individual study, the funnel plot became more symmetrical. Clearly, there may be many different interpretations about an asymmetrical funnel plot.[44] Strictly speaking, the funnel plot and related statistical techniques are useful to test the existence of so-called "small study bias",[326] although publication bias is one of many possible interpretations of the observed association between effect size and sample size in meta-analysis.

Implications for decision makers and researchers

The potential problem of publication and other selection biases should be taken into consideration by all who are involved in evidence-based decision making. For research funding bodies and RECs, all funded or approved studies should be registered prospectively and such registrations should be accessible to the public. The dissemination of research results should be considered as an integral part of research.

In all systematic reviews, a thorough literature search is crucial to identifying all relevant studies, published or not. Whenever possible, registers of clinical trials should be used in a systematic review to identify relevant studies and identified trials should be classified according to whether they are prospectively or retrospectively registered. A search of electronic databases alone is seldom sufficient and should be supplemented by checking the references of relevant studies and contacting experts or organisations.[275]

Although it is still controversial, systematic reviews should not routinely exclude unpublished studies or abstracts presented at scientific meetings. The quality of unpublished studies or abstracts should be assessed using the same criteria as for published studies. Sensitivity analysis could be used in systematic reviews to

compare the results with and without data from unpublished studies.

Issues such as selective reporting of multiple outcomes, duplicate publication and language of publication should also be considered in systematic reviews. Bias may occur in systematic reviews when relevant studies are not included owing to a lack of data on relevant outcomes. Duplicate publication should be carefully identified and excluded. Systematic reviews should not include or exclude studies based on the languages in which they are published. When language restriction has to be used, its possible impact on the results of a systematic review should be assessed.

Even a thorough literature search cannot eliminate publication bias; therefore the risk should be assessed and the estimated risk of publication bias should be incorporated into the review's conclusions and recommendations. The risk of publication bias can be assessed, for example, according to the sample sizes used, the potential number of studies that may have been conducted, research sponsors, times of publication, and heterogeneity across studies.

Some statistical and modelling methods are available to deal with publication bias in meta-analyses, such as: the fail-safe N method; funnel plots; and rank correlation, linear regression, trim and fill, and some complex modelling methods. The fail-safe N and funnel plot-related statistical methods are the most commonly used. These methods are mainly useful for detecting publication bias, although some methods (e.g. trim and fill) could provide an estimate by adjusting for the detected publication bias. It should be stressed that all these methods are by nature indirect and exploratory, because the true extent of publication bias is generally unknown.

There are some methodological difficulties in using the available methods to assess publication bias in meta-analyses. In most cases it is impossible to separate the influence of factors other than publication bias on the observed association between the estimated effects and sample sizes across studies. The appropriateness of many methods is based on some strict assumptions that can be difficult to justify in practice (see *Table 8*). For these reasons, it seems reasonable to argue that these methods "are not very good remedies for publication bias".[16] The attempt at identifying or adjusting for publication bias in a meta-analysis should be used mainly for the purpose of sensitivity analyses, and the results should be interpreted with great caution.

Large-scale confirmatory trials become necessary after a systematic review has reported a clinically significant finding but publication bias cannot be safely excluded as an alternative explanation. As compared with a universal register of all trials, confirmatory large trials are more selective about the research areas and objectives, but at the same time more flexible in their ability to minimise the impact of other biases, for example, biases related to study design, selection of a control, participants and setting.

Recommendations for future research

- Further research is needed to provide more direct empirical evidence about publication and related biases. There is a lack of direct comparison between published studies and studies in an unbiased sampling frame. In particular, there is a lack of evidence about the impact of publication bias on health decision making and the outcomes of patient management.
- In the foreseeable future, many systematic reviews may still have to depend upon studies identified retrospectively from the published literature. The available methods for dealing with publication bias should be evaluated by comparing their assumptions, performance and results, ideally by using a set of meta-analyses in which the extent of publication bias could be estimated according to unbiased samples of relevant studies.
- Research is also needed to develop new methods that are robust and easy to use for detecting publication bias in systematic reviews. In particular, there is a lack of methods that can be used to detect publication bias in narrative systematic reviews.
- It is most important for future research to identify feasible and efficient methods for preventing publication bias. Further research is needed to answer questions about: how to establish and maintain the prospective registration of clinical trials and observational studies; how to make all research findings accessible to the public; and how the developments in computer science and information technology can be used to solve the problem of publication bias.

- The problem of publication bias is unlikely to be solved by the conventional paper-based medical journals because of their intrinsic limitations. Further research concerning publication bias should be an integral part of investigation that explores alternatives to the conventional methods for generating, disseminating, preserving and utilising scientific research findings.

Acknowledgements

The work was commissioned by the NHS R&D Health Technology Assessment programme. The authors would like to thank Kathleen Wright for assisting with the searching and location of the literature, and Sue Duval at the University of Colorado Health Sciences Centre for help in using the trim and fill method. We thank Iain Chalmers, Trevor Sheldon, Jos Kleijnen, Andrew Stevens and Keith Abrams for commenting on earlier manuscripts. We also thank the Health Technology Assessment referees who provided very helpful comments.

References

1. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;**309**:597–9.
2. Light RJ, Pillemer DB. Summing up: the science of reviewing research. Cambridge, MA: Harvard University Press, 1984.
3. Chalmers I, Altman DG, editors. Systematic reviews. London: BMJ Publishing Group, 1995.
4. Felson DT. Bias in meta analytic research. *J Clin Epidemiol* 1992;**45**:885–92.
5. Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979; **86**:638–41.
6. Boissel JP, Haugh MC. The iceberg phenomenon and publication bias – the editor's fault? *Clin Trials Meta-Anal* 1993;**28**:309–15.
7. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc A* 1988; **151**:419–63.
8. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Am Stat Assoc J* 1959;**54**:30–4.
9. Chalmers TC, Grady GF. A note on fatality in serum hepatitis. *Gastroenterology* 1965;**49**:22–6.
10. Boisen E. Testicular size and shape of 47,XYY and 47,XXY men in a double-blind, double-matched population survey. *Am J Hum Genet* 1979;**31**:697–703.
11. Simes RJ. Confronting publication bias: a cohort design for meta analysis. *Stat Med* 1987;**6**:11–29.
12. Song F, Gilbody S. Increase in studies of publication bias coincided with increasing use of meta-analysis [Letter]. *BMJ* 1998;**316**:471.
13. Last JM. A dictionary of epidemiology. 3rd ed. New York: Oxford University Press, 1995.
14. Everitt BS. The Cambridge dictionary of statistics in the medical sciences. Cambridge: Cambridge University Press, 1995.
15. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990;**263**:1392–5.
16. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**:1385–9.
17. Dickersin K. Confusion about "negative" studies [Letter]. *N Engl J Med* 1990;**322**:1084.
18. Smith R. What is publication [Editorial]? *BMJ* 1999;**318**:142.
19. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;**32**:51–3.
20. Rosenthal R. Publication bias, retrieval bias and pipeline effects. Discussion of the paper by Begg and Berlin. *J R Stat Soc A* 1988;**151**:419–63.
21. Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. *JAMA* 1998;**279**:319–20.
22. Ben-Shlomo Y, Davey-Smith G. "Place of publication" bias [Letter]? *BMJ* 1994;**309**:274.
23. Tramer MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;**315**:635–40.
24. Gregoire G, Derderian F, Lorier JL. Selecting the language of the publications included in a meta-analysis: is there a tower of babel bias? *J Clin Epidemiol* 1995;**48**:159–63.
25. Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;**295**:654–6.
26. Centre for Reviews and Dissemination, University of York. http://www.york.ac.uk/inst/crd
27. Sohn D. Publications bias and the evaluation of psychotherapy efficacy in reviews of the research literature. *Clin Psychol Rev* 1996;**16**:147–56.
28. Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited – the effect of the outcome of statistical tests on the decision to publish and vice-versa. *Am Stat* 1995;**49**:108–12.
29. Smart RG. The importance of negative results in psychological research. *Can Psychol* 1964;**5**:225–32.
30. Bozarth JD, Roberts RR. Signifiying significant significance. *Am Psychol* 1972;**27**:774–5.
31. Hubbard R, Armstrong JS. Publication bias against null results. *Psychol Rep* 1997;**80**:337–8.
32. Greenwald AG. Consequences of prejudice against the null hypothesis. *Psychol Bull* 1975;**82**:1–20.
33. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;**1**:155–8.
34. Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med* 1994;**1**:267–71.

35. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;**272**:122–4.

36. Mulward S, Gotzsche PC. Sample-size of randomized double-blind trials 1976–1991. *Dan Med Bull* 1996;**43**:96–8.

37. Csada RD, James PC, Espie RHM. The file drawer problem of nonsignificant results – does it apply to biological research? *Oikos* 1996;**76**:591–3.

38. Bauchau V. Is there a "file drawer problem" in biological research? *Oikos* 1997;**79**:407–9.

39. Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989;**84**:381–92.

40. McPherson K. Discussion of the paper by Begg and Berlin: Publication bias: a problem in interpreting medical data. *J R Stat Soc A* 1988;**151**:445–7.

41. Allison DB, Faith MS, Gorman BS. Publication bias in obesity treatment trials. *Int J Obes* 1996;**20**:931–7.

42. Egger M, Davey-Smith G. Misleading meta-analysis: lessons from "an effective, safe, simple" intervention that wasn't. *BMJ* 1995;**310**:752–4.

43. Pipkin F, Crowther C, de-Swiet M, Duley L, Judd A, Lilford R, *et al.* Where next for prophylaxis against pre-eclampsia? *Br J Obstet Gynaecol* 1996;**103**:603–7.

44. Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.

45. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Control Clin Trials* 1987;**8**:343–53.

46. Rotton J, Foos PW, Vanmeek L, Levitt M. Publication practices and the file drawer problem – a survey of published authors. *J Soc Behav Pers* 1995;**10**:1–13.

47. Coursol A, Wagner EE. Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias. *Prof Psychol* 1986;**17**:136–7.

48. Sommer B. The file drawer effect and publication rates in menstrual cycle research. *Psychol Wom Q* 1987;**11**:233–42.

49. Shadish WR, Doherty M, Montgomery LM. How many studies are in the file drawer – an estimate from the family marital psychotherapy literature. *Clin Psychol Rev* 1989;**9**:589–603.

50. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;**84**:374–80.

51. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**:867–72.

52. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow up of applications submitted to two institutional review boards. *JAMA* 1992;**267**:374–8.

53. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993; Doc. no. 50.

54. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;**315**:640–5.

55. Dickersin K. How important is publication bias? A synthesis of available data. *Aids Educ Prev* 1997;**9**:15–21.

56. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;**4**:1529–41.

57. Ioannidis J. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;**279**:281–6.

58. Misakian AL, Bero LA. Publication bias and research on passive smoking. Comparison of published and unpublished studies. *JAMA* 1998;**280**:250–3.

59. Rothwell PM, Robertson G. Meta-analyses of randomised controlled trials. *Lancet* 1997; **350**:1181–2.

60. Auger CP. Information sources in grey literature. 4th ed. (Guides to information sources.) London: Bowker Saur, 1998.

61. Smith ML. Publication bias and meta-analysis. *Eval Educ* 1980;**4**:22–4.

62. Glass GV, McGaw B, Smith ML. Meta-analysis in social research. London: Sage, 1981.

63. White KR. The relation between socioeconomic status and academic achievement. *Psychol Bull* 1982;**91**:461–81.

64. Detsky A, Baker J, O'Rourke K, Goel V. Perioperative parenteral nutrition: a meta-analysis. *Ann Intern Med* 1987;**107**:195–203.

65. Devine EC. Empirical assessment of publication bias: lessons from two meta-analyses [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:60.

66. MacLean CH, Morton SC, Straus WL, Ofman J, Roth E, Shekelle PG. Unpublished data from United States Food and Drug Administration New Drug Application Reviews: how do they compare to published data when assessing nonsteroidal antiinflammatory drug (NSAID) associated dyspepsia [Abstract]? In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:42.

67. McAuley L, Moher D, Pham B, Tugwell P. Evaluation of the impact of grey literature on meta-analysis [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:17.

68. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;**272**:158–62.

69. DeBellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Ann Oncol* 1992;**3**:187–91.

70. Chalmers I, Adams M, Dickersin K, Hetherington J, Tamow-Mordi W, Meinert C, *et al.* A cohort study of summary reports of controlled trials. *JAMA* 1990;**263**:1401–5.

71. Loep M, Kleijnen J. Full publication of abstracts initially published in the *Netherlands Journal of Medicine*. Unpublished manuscript, 1999.

72. Landry VL. The publication outcome for the papers presented at the 1990 ABA conference. *J Burn Care Rehabil* 1996;**17**:23A–6A.

73. Cheng K, Preston C, Ashby D, OHea U, Smyth R. Time to publication as full reports of abstracts of randomized controlled trials in cystic fibrosis. *Pediatr Pulmonol* 1998;**26**:101–5.

74. Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting. Why investigators fail to publish. *JAMA* 1998;**280**:257–9.

75. Petticrew M, Gilbody S, Song F. Lost information: the fate of papers presented at the 40th Society for Social Medicine Conference. *J Epidemiol Community Health* 1999;**53**:442–3.

76. Goldman L, Loscalzo A. Fate of cardiology research originally published in abstract form. *N Engl J Med* 1980;**303**:255–8.

77. Corry AM. A survey of the publication history of randomly selected IADR/AADR abstracts presented in 1983 and 1984. *J Dent Res* 1990;**69**:1453–5.

78. Juzych M, Shin DH, Coffey JB, Parrow KA, Tsai CS, Briggs KS. Pattern of publication of ophthalmic abstracts in peer-reviewed journals. *Ophthalmology* 1991;**98**:553–6.

79. McCormick MC, Holmes JH. Publication of research presented at the pediatric meetings. Change in selection. *Am J Dis Child* 1985;**139**:122–6.

80. Meranze J, Ellison N, Greenhow DE. Publications resulting from anesthesia meeting abstracts. *Anesth Analg* 1982;**61**:445–8.

81. Maxwell MB. Published or perished: what becomes of papers presented at Oncology Nursing Society Congresses? *Oncol Nurs Forum* 1981;**8**:73–4.

82. Dudley HAF. Surgical research: master of servant. *Am J Surg* 1978;**135**:458–60.

83. Agustsdottir A, Holcombe J, Wright P, Daffin P, Ogletree G. Publication of patient-related oncology nursing research. *Oncol Nurs Forum* 1995;**22**:827–30.

84. Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. *Can J Anaesth* 1993;**40**:632–4.

85. Elder NC, Blake RL. Publication patterns of presentations at the Society of Teachers of Family Medicine and North American Primary Care Research Group annual meetings. *Fam Med* 1994;**26**:352–5.

86. Gavazza JB, Foulkes GD, Meals RA. Publication pattern of papers presented at the American Society for Surgery of the Hand annual. *J Hand Surg Am* 1996;**21**:742–5.

87. Koren G, Graham K, Shear H, Einarson T. Bias against the null hypothesis: the reproductive hazards of cocaine. *Lancet* 1989;**ii**:1440–2.

88. Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998;**280**:254–7.

89. Chokkalingam A, Scherer R, Dickersin K. Agreement of data in abstracts compared to full publications. *Control Clin Trials* 1998;**19**:61S–2S.

90. Ohlsson A, Walia R. Differences between information provided in abstracts of randomized controlled trials in neonates submitted to the annual meeting of the American Pediatric Society and the Society for Pediatric Research (APS/SPR) and final publications – implications for meta-analyses [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:19.

91. Bero LA, Galbraith A, Rennie D. Sponsored symposia on environmental tobacco smoke. *JAMA* 1994;**271**:612–17.

92. West RR, Jones DA. Publication bias in statistical overview of trials: example of psychological rehabilitation following myocardial infarction [Abstract]. In: Proceedings of the 2nd International Conference on the Scientific Basis of Health Services and 5th Annual Cochrane Colloquium; 1997 Oct 8–12; Amsterdam, 82.

93. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;**317**:426–32.

94. Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996; **88**:206–7.

95. Jefferson T. Redundant publication in biomedical sciences: scientific misconduct or necessity? *Sci Eng* 1998;**4**:135–40.

96. Lock S. Repetitive publication: a waste that must stop. *BMJ* 1984;**288**:661–2.

97. Curfman GD. Duplicate publication on post-menopausal bone loss. *N Engl J Med* 1987; **317**:833–4.

98. Notice of duplicate publication [Editorial]. *JAMA* 1985;**254**:3423.

99. Fulginiti VA. Unfortunately, more on duplicate publication. *Am J Dis Child* 1985;**139**:865–6.

100. Angell M, Relman AS. Redundant publication. *N Engl J Med* 1989;**320**:1212–13.

101. Sole contribution [Editorial]. *N Engl J Med* 1969; **281**:676–7.

102. Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;**36**:429–32.

103. Vandekerckhove P, O'Donovan P, Lilford R, Harada T. Infertility treatment: from cookery to science. The epidemiology of randomised controlled trials. *Br J Obstet Gynaecol* 1993;**100**:1005–36.

104. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996; **347**:1024–6.

105. Vandenbrouche JP. On not being born a native speaker of English. *BMJ* 1989;**298**:1461–2.

106. Bakewell D. Publish in English, or perish [Letter]? *Nature* 1992;**356**:648.

107. Henrissat B. National publication bias [Letter]. *Nature* 1991;**354**:427.

108. Braun T, Schubert A. National publication bias. *J Inform Sci* 1993;**19**:75–6.

109. Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts. Effects of referee characteristics and publication language. *JAMA* 1994;**272**:149–51.

110. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–9.

111. Ottenbacher K, Difabio RP. Efficacy of spinal manipulation/mobilization therapy. A meta-analysis. *Spine* 1985;**10**:833–7.

112. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;**19**:159–66.

113. Tang JL, Zhan SY, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;**319**:160–1.

114. Egger M, Davey-Smith G. Bias in location and selection of studies. *BMJ* 1998;**316**:61–6.

115. Zielinski C. New equities of information in an electronic age. The third world needs first world information – how about the other way round? *BMJ* 1995;**310**:1480–1.

116. Nieminen P, Isohanni M. Bias against European journals in medical publication databases [Letter]. *Lancet* 1999;**353**:1592.

117. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Control Clin Trials* 1985;**6**:306–17.

118. Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus MEDLARS searches. *Control Clin Trials* 1988;**9**:23–31.

119. Kirpalani H, Schmidt B, McKibbon KA, Hayes RB, Sinclair JC. Searching MEDLINE for randomised clinical trials involving care of the newborn. *Pediatrics* 1989;**83**:543–6.

120. Gotzsche PC, Lange B. Comparison of search strategies for recalling double-blind trials from Medline. *Dan Med Bull* 1991;**38**:476–8.

121. Kleijnen J, Knipschild P. The comprehensiveness of MEDLINE and EMBASE computer searches. Searches for controlled trials of homoeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. *Pharm Weekbl [Sci]* 1992;**14**:316–20.

122. Silagy CA. Developing a register of randomised controlled trials in primary care. *BMJ* 1993; **306**:897–900.

123. Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993; Doc. no. 33.

124. Adams CE, Power A, Frederick K, Lefebvre C. An investigation of the adequacy of Medline searches for randomized controlled trials (RCTs) on the effects of mental health care. *Psychol Med* 1994; **24**:741–8.

125. Taus C, Pucci E, Giuliani G, Telaro E, Pistotti V. The use of "grey literature" in a sub-set of neurological Cochrane reviews [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:30.

126. McDonald S, Taylor L, Adams C. Searching the right database. A comparison of four databases for psychiatry journals. *Health Libr Rev* 1999;**16**:151–6.

127. Brooks TA. Private acts and public objects: an investigation of citer motivations. *J Am Soc Inf Sci* 1985;**36**:223–9.

128. Shadish WR, Tolliver D, Gray M, SenGupta SK. Author judgements about works they cite: three studies from psychology journals. *Soc Stud Sci* 1995;**25**:477–98.

129. Christensenszalanski JJJ, Beach LR. The citation bias – fad and fashion in the judgment and decision literature. *Am Psychol* 1984;**39**:75–8.

130. Robins RW, Craik KH. Is there a citation bias in the judgment and decision literature? *Organ Behav Hum Decis Process* 1993;**54**:225–44.

131. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;**305**:15–19.

132. Ravnskov U. Quotation bias in reviews of the diet–heart idea. *J Clin Epidemiol* 1995;**48**:713–9.

133. Hutchison BG, Oxman AD, Lloyd S. Comprehensiveness and bias in reporting clinical trials. Study of reviews of pneumococcal vaccine effectiveness. *Can Fam Physician* 1995;**41**:1356–60.

134. Song F, Landes D, Glenny A, Sheldon T. Prophylactic removal of impacted third molars: an assessment of published reviews. *Br Dent J* 1997;**182**:339–46.

135. Joyce J, Rabe-Hesketh S, Wessely S. Reviewing the reviews. The example of chronic fatigue syndrome. *JAMA* 1998;**280**:264–6.

136. Budd JM, Sievert M, Schultz TR. Phenomena of retraction. Reasons for retraction and citations to the publications. *JAMA* 1998;**280**:296–7.

137. de Semir V. What is newsworthy? *Lancet* 1996; **347**:1163–6.

138. Combs B, Slovic P. Causes of death: biased newspaper coverage and biased judgments. *Journalism Q* 1979;**56**:837–43.

139. Houn F, Bober MA, Huerta EE, Hursting SD, Lemon S, Weed DL. The association between alcohol and breast cancer popular press coverage of research. *Am J Public Health* 1995;**85**:1082–6.

140. Koren G, Klein N. Bias against negative studies in newspaper reports of medical research. *JAMA* 1991;**266**:1824–6.

141. Wing S, Shy C, Wood J, Wolf S, Cragle D, Frome E. Mortality among workers at Oak Ridge National Laboratory. *JAMA* 1991;**265**:1397–402.

142. Jablon S, Hrubec Z, Boice JJ. Cancer in populations living near nuclear facilities: a survey of mortality nationwide and incidence in two states. *JAMA* 1991;**265**:1403–8.

143. Chalmers I. Under-reporting research is scientific misconduct. *JAMA* 1990;**263**:1405–8.

144. Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction – an example of publication bias. *Int J Cardiol* 1993;**40**:161–6.

145. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomised trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; **321**:406–12.

146. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricisine on survival after myocardial infarction. *N Engl J Med* 1992;**327**:227–33.

147. Yamey G. Scientists who do not publish trial results are "unethical" [News]. *BMJ* 1999;**319**:939.

148. Yusuf S, Koon T, Woods K. Intravenous magnesium in acute myocardial infarction: an effective, safe, simple, and inexpensive intervention. *Circulation* 1993;**87**:2043–6.

149. ISIS-4 Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58050 patients with acute myocardial infarction. *Lancet* 1995;**345**:669–85.

150. Government Statistical Service. Finished consultant episodes by diagnosis, operation and specialty. England: financial year 1993–94 (Hospital episode statistics, vol. 1). London: GSS, 1995.

151. Skolbekken JA. The risk epidemic in medical journals. *Soc Sci Med* 1995;**40**:291–305.

152. Angell M, Kassirer JP. Clinical research – what should the public believe? *N Engl J Med* 1994; **331**:189–90.

153. Taubes G. Epidemiology faces its limits. *Science* 1995;**269**:164–9.

154. MacCoun R. Biases in the interpretation and use of research results. *Annu Rev Psychol* 1998;**49**:259–87.

155. Campillo C. Publication bias in two Spanish medical journals. (The International Congress on Biomedical Peer review and Global Communications. Prague, 1997.) In: The Cochrane Review Methodology Database; The Cochrane Library (issue 1). Oxford: Update Software, 2000.

156. Rothwell PM, Slattery J, Warlow CP. A systematic review of the risks of stroke and death due to endarterectomy for symptomatic carotid stenosis. *Stroke* 1996;**27**:260–5.

157. Frank E. Authors' criteria for selecting journals. *JAMA* 1994;**272**:163–4.

158. Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS. Withholding research results in academic life science. *JAMA* 1997;**277**:1224–8.

159. Kupfersmid J, Fiala M. A survey of attitudes and behaviors of authors who publish in psychology and education journals. *Am Psychol* 1991;**46**:249–50.

160. Kerr S, Tolliver J, Petree D. Manuscript characteristics which influence acceptance for management and social science journals. *Acad Manage J* 1977; **20**:132–41.

161. View. *BMJ* 1983;**287**:1886.

162. Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA* 1994;**272**:96–7.

163. *The Lancet*. http://www.thelancet.com

164. Abby M, Massey MD, Galandiuk S, Polk HC. Peer review is an effective screening process to evaluate medical manuscripts. *JAMA* 1994;**272**:105–7.

165. Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983;**1**:164–9.

166. Manuscript guidelines [Editorial]. *Diabetologia* 1984;**25**:4A.

167. Angell M. Negative studies. *N Engl J Med* 1989; **321**:464–6.

168. Bailar JC, Patterson K. Journal peer review: the need for a research agenda. *N Engl J Med* 1985;**312**:654–7.

169. Godlee F, Dickersin K. Bias, subjectivity, chance, and conflict of interest in editorial decisions. In: Godlee F, Jefferson T, editors. Peer review in health sciences. London: BMJ Books, 1999:57–78.

170. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Ther Res* 1977;**1**:161–75.

171. Ernst E, Resch KL. Reviewer bias – a blinded experimental study. *J Lab Clin Med* 1994;**124**:178–82.

172. Epstein WM. Confirmation response bias among social work journals. *Sci Techol Hum Values* 1990;**15**:9–38.

173. Abbot NC, Ernst E. Publication bias: direction of outcome less important than scientific quality. *Perfusion* 1998;**11**:182–4.

174. Resch KL, Ernst E, Garrow J. Does peer review favor the conceptual framework of orthodox medicine? A randomized controlled study. (The International Congress on Biomedical Peer Review and Global Communications. Prague, 1997.) In: The Cochrane Review Methodology Database; The Cochrane Library (issue 1). Oxford: Update Software, 2000.

175. Ernst E, Resch KL. Peer-reviewer bias against unconventional medicine? (The International Congress on Biomedical Peer Review and Global Communications. Prague, 1997.) In: The Cochrane Review Methodology Database; The Cochrane Library (issue 1). Oxford: Update Software, 2000.

176. Garfunkel JM, Ulshen MH, Hamrick HJ, Lawson EE. Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA* 1994;**272**:137–8.

177. Ector H, Aubert A, Stroobandt R. Review of the reviewer. *Pacing Clin Electrophysiol* 1995;**18**:1215–17.

178. Justice AC, Berlin JA, Fletcher SW, Fletcher RH, Goodman SN. Do readers and peer reviewers agree on manuscript quality? *JAMA* 1994;**272**:117–19.

179. Blumenthal D, Causino N, Campbell E, Seashore K. Relationships between academic institutions and industry in the life sciences – an industry survey. *N Engl J Med* 1996;**334**:368–73.

180. Blumenthal D, Gluck M, Louis K, Wise D. Industrial support of university research in biotechnology. *Science* 1986;**231**:242–6.

181. Blumenthal D, Campbell EG, Causino N, Seashore K. Participation of life-science faculty in research relationships with industry. *N Engl J Med* 1996;**335**:1734–9.

182. Rosenberg SA. Secrecy in medical research. *N Engl J Med* 1996;**334**:392–4.

183. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;**280**:833–6.

184. Bardy AH. Report bias in drug research. *Therapie* 1996;**51**:382–3.

185. Bardy A. Bias in reporting clinical trials. *Br J Clin Pharmacol* 1998;**46**:147–50.

186. Dent THS, Hawke S. Too soon to market: doctors and patients need more information before drugs enter routine use. *BMJ* 1997;**315**:1248–9.

187. Abraham J, Lewis G. Secrecy and transparency of medicines licensing in the EU. *Lancet* 1998; **352**:480–2.

188. Roberts I, Li-Wan-Po A, Chalmers I. Intellectual property, drug licensing, freedom of information, and public health. *Lancet* 1998;**352**:726–9.

189. Rennie D. Thyroid storm. *JAMA* 1997;**277**:1238–43.

190. Lauritsen K, Kavelund T, Larsen LS, Rask-Madsen J. Withholding unfavourable results in drug company sponsored clinical trials [Letter]. *Lancet* 1987;**i**:1091.

191. Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL, *et al.* Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JAMA* 1997; **277**:1205–13.

192. Canadian Coordinating Office for Health Technology Assessment. Annual report 1997–1998. Ottawa: CCOHTA, 1998.

193. Millstone E, Brunner E, White I. Plagiarism or protecting public health. *Nature* 1994;**371**:647–8.

194. Rochon PA, Gurwitz JH, Simms RW, Fortin PR, Felson DT, Minaker KL, *et al.* A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Intern Med* 1994;**154**:157–63.

195. Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med* 1998;**338**:101–6.

196. Meltzer JI. Conflict of interest in the debate over calcium-channel antagonists [Letter]. *N Engl J Med* 1998;**338**:1696.

197. Opie L. Conflict of interest in the debate over calcium-channel antagonists [Letter]. *N Engl J Med* 1998;**338**:1696–7.

198. Strandgaard S. Conflict of interest in the debate over calcium-channel antagonists [Letter]. *N Engl J Med* 1998;**338**:1697.

199. Detsky AS, Stelfox HT. Conflict of interest in the debate over calcium-channel antagonists [Letter]. *N Engl J Med* 1998;**338**:1698.

200. Smith R. Beyond conflict of interest: transparency is the key. *BMJ* 1998;**317**:291–2.

201. Barnes DE, Bero LA. Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law* 1996;**21**:515–42.

202. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996;**124**:485–9.

203. Lane DM, Dunlap WP. Estimating effect size: bias resulting from the significance criterion in editorial decisions. *Br J Math Stat Psychol* 1978;**31**:107–12.

204. Hedges LV. Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistically insignificant mean differences. *J Educ Stat* 1984;**9**:61–85.

205. Begg CB. Publication bias. In: Cooper H, Hedge LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation, 1994:399–409.

206. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;**81**:107–15.

207. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

208. Fox T. Crisis in communication: the functions and future of medical journals. London: Athlone Press, 1965.

209. Newcombe RG. Discussion of the paper by Begg and Berlin: Publication bias: a problem in interpreting medical data. *J R Stat Soc A* 1988;**151**:448–9.

210. Julian D. Meta-analysis and the meta-epidemiology of clinical research. Registration of trials should be required by editors and registering agencies [Letter]. *BMJ* 1998;**316**:311.

211. Smith R, Roberts I. An amnesty for unpublished trials: send us details on any unreported trials [Editorial]. *BMJ* 1997;**315**:622.

212. Horton R. Medical editors' trial amnesty [Editorial]. *Lancet* 1997;**350**:756.

213. Cochrane Controlled Trials Register. In: The Cochrane Library (issue 4). Oxford: Update Software, 1999.

214. O'Toole L. Too many hoops? *MRC News* 1998;**79**:1.

215. McDonald S, Clarke M. 50 years of reporting of randomised controlled trials in general health care journals worldwide: how things have changed [Abstract]. In: proceedings of the 6th International Cochrane Colloquium; 1998 Oct 22–26; Baltimore, MD: 52.

216. Fletcher RH, Fletcher SW. The effectiveness of editorial peer review. In: Godlee F, Jefferson T, editors. Peer review in health sciences. London: BMJ Books, 1999:45–56.

217. Pierie JP, Walvoort HC, Overbeke AJ. Readers' evaluation of effect of peer review and editing on quality of articles in the *Nederlands Tijdschrift voor Geneeskunde*. *Lancet* 1996;**348**:1480–3.

218. Smith R. The future of peer review. In: Godlee F, Jefferson T, editors. Peer review in health sciences. London: BMJ Books, 1999:244–53.

219. Laband DN, Piette MJ. A citation analysis of the impact of blinded peer review. *JAMA* 1994; **272**:147–9.

220. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. *JAMA* 1998;**280**:237–40.

221. Rooyen SV, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review. A randomized trial. *JAMA* 1998; **280**:234–7.

222. Yankauer A. How blind is blind review? *Am J Public Health* 1991;**81**:843–5.

223. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review: a randomized trial. *JAMA* 1990;**263**:1371–6.

224. Cho MK, Justice AC, Winker MA, Berlin JA, Waeckerle JF, Callaham ML, *et al.* Masking author identity in peer review. What factors influence masking success? *JAMA* 1998;**280**:243–5.

225. Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D. Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA* 1998;**280**:240–2.

226. Rennie D. Freedom and responsibility in medical publication. Setting the balance right. *JAMA* 1998;**280**:300–2.

227. Smith R. Opening up BMJ peer review: a beginning that should lead to complete transparency. *BMJ* 1999;**318**:4–5.

228. Huth EJ. Electronic publishing in the health sciences. *Bull Pan Am Health Organ* 1995;**29**:81–7.

229. Bero L. The electronic future: what might an online scientific paper look like in five years' time [Editorial]? *BMJ* 1997;**315**:1692.

230. Delamothe T. The electronic future of scientific articles. *Lancet* 1998;**351**:5–6.

231. Fletcher RH, Fletcher SW. The future of medical journals in the western world. *Lancet* 1998;**352**:30–3.

232. Varmus H. E-Biomed: a proposal for electronic publications in the biomedical sciences. In: www.thelancet.com/newlancet/sub/nochange/body.nihdiss_1.html Bethesda, MD: National Institutes of Health, 1999.

233. Cameron G. Electronic database and the scientific record. In: Butterworth I, editor. The impact of electronic publishing on the academic community. An international workshop organised by the Academia Europaea and the Wenner-Gren Foundation. 1997 Apr 16–20; Stockholm. http://tiepac.portlandpress.co.uk London: Portland Press, 1997.

234. Sim I, Rennels G. A trial bank model for the publication of clinical trials. In: Proceedings of the Annual Symposium on Computer Applications in Medical Care; 1995 Nov 5–9; New Orleans. New Orleans: American Medical Informatics Association, 1995:863–7.

235. Delamothe T, Mullner M, Smith R. Pleasing both authors and readers: a combination of short print articles and longer electronic ones may help us do this. *BMJ* 1999;**318**:888–9.

236. Chalmers I, Altman DG. How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 1998;**353**:490–3.

237. Berlin JA. Will publication bias vanish in the age of online journals? *Online J Curr Clin Trials* 1992; (Jul 8): Doc. no. 12.

238. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research: meta-analysis is an important contribution to research and practice but it's not a panacea. *BMJ* 1997;**315**:617–19.

239. Song F, Eastwood A, Gilbody S, Duley L. The role of electronic journals in reducing publication bias. *Med Inf* 1999;**24**:223–9.

240. Current Controlled Trials. Ways and means of registering controlled trials. In: The metaRegister of Controlled Trials (mRCT), 1999. http://www.controlled-trials.com

241. Boissel JP, Ad Hoc Working Party of the International Collaborative Group on Clinical Trials Registries. Position paper and consensus recommendation on clinical trial registries. *Clin Trials MetaAnal* 1993;**28**:255–66.

242. Dickersin K. Keeping posted. Why register clinical trials? – revisited. *Control Clin Trials* 1992;**13**:170–7.

243. Easterbrook PJ. Directory of registries of clinical trials. *Stat Med* 1992;**11**:345–423.

244. Fayers P, Gibson D, Mossman J. UKCCCR Register of UK cancer trials. *Control Clin Trials* 1995; **16**:172–81.

245. Langhorne P. Prospectively identified trials could be used for comparison with meta-analyses [Letter]. *BMJ* 1998;**316**:471.

246. Kleijnen J, Knipschild P. Review articles and publication bias. *Arzneimittelforschung/Drug Res* 1992;**42**:587–91.

247. Verstraete M. Registry of prospective clinical trials second report. *Thromb Haemost* 1976;**36**:239–50.

248. The Ottawa Stroke Trials Registry Collaborative Group and the Development of the Ottawa Stroke Trials Registry (OSTR). *Control Clin Trials* 1994;**15**:503–11.

249. Cochrane Collaborative Review Groups. The Cochrane database of systematic reviews. In: The Cochrane Library (issue 30). Oxford: Update Software, 1998.

250. Robertson SE, Mayans MV, Horsfall S, Wright PF, Clemens J, Ivanoff B, *et al.* The WHO global programme for vaccines and immunization vaccine trial registry. *Bull World Health Organ* 1997;**75**:295–305.

251. National Research Register. In: http://www.doh.gov.uk/research/nrr.htm Oxford: Update Software, 1999.

252. Downs JR, Gotto A, Clearfield M, Gordon D, Manolio T, Goldbourt U, *et al.* Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. *Am J Cardiol* 1995;**75**:1130–4.

253. Simes RJ. Prospective meta-analysis of cholesterol-lowering studies – the prospective pravastatin pooling (Ppp) project and the Cholesterol Treatment Trialists (Ctt) Collaboration. *Am J Cardiol* 1995;**76**:C122–6.

254. The metaRegister of Controlled Trials (mRCT). In: http://www.controlled-trials-com London: Current Controlled Trials, 1999.

255. Meinert CL. Toward prospective registration of clinical trials. *Control Clin Trials* 1988;**9**:1–5.

256. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996;**313**:1390–3.

257. Dickersin K, Garcia Lopez F. Regulatory process effects clinical trial registration in Spain. *Control Clin Trials* 1992;**13**:507–12.

258. Chollar S. A registry for clinical trials. *Ann Intern Med* 1998;**128**:701–2.

259. Tonks A. Registering clinical trials. *BMJ* 1999;**319**: 1565–8.

260. Dickersin K. Report from the panel on the case for registers of clinical trials at the eighth annual meeting of the Society for Clinical Trials. *Control Clin Trials* 1988;**9**:76–81.

261. Bossard N, Boissel JP. Registry of multicenter clinical trials: 14th and 15th report 1992–1993. *Control Clin Trials* 1996;**17**:130–75.

262. Dickersin K. International collaboration in trial registration [Oral presentation]. In: Registering information about randomised controlled trials. Meeting organised by the BMJ Publishing Group, *The Lancet*, and the Association of the British Pharmaceutical Industry; 1999 Oct 5; London.

263. Freedom of information, when it suits [Editorial]. *Lancet* 1998;**352**:665.

264. Bardy AH. Freedom of information [Letter]. *Lancet* 1998;**352**:1229–30.

265. Eckert CH. Bioequivalence of levothyroxine preparations: industry sponsorship and academic freedom [Letter]. *JAMA* 1997;**277**:1200–201.

266. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;**48**:23–40.

267. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987;**6**:315–25.

268. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995;**345**:772–6.

269. Cappelleri JC, Ioannidis JPA, Schmid CH, Ferranti SDD, Aubert M, Chalmers TC, *et al.* Large trials vs meta-analysis of smaller trials. How do their results compare? *JAMA* 1996;**276**:1332–8.

270. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;**337**:536–42.

271. Ioannidis J, Cappelleri J, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA* 1998;**279**:1089–93.

272. Flournoy N, Olkin I. Do small trials square with large ones? *Lancet* 1995;**345**:741–2.

273. Odaka T, Nakayama A, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T, *et al.* The effect of a multiple literature database search – a numerical evaluation in the domain of Japanese life science. *J Med Syst* 1992;**16**:177–81.

274. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M. Using the Cochrane Controlled Clinical Trials Register (CENTRAL) to identify clinical trials in selected medical fields [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:29.

275. McManus R, Wilson S, Delaney B, Fitzmaurice D, Hyde C, Tobias R, *et al.* Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *BMJ* 1998;**317**:1562–3.

276. Topfer L, Parada A, Menon D, Noorani H, Perras C, Serra-Prat M. Comparison of literature searches on quality and costs for health technology assessment using the Medline and Embase databases. *Int J Technol Assess Health Care* 1999;**15**:297–303.

277. Jadad AR, Moher D, Klassen TP. Guides for reading and interpreting systematic reviews. II: How did the authors find the studies and assess their quality? *Arch Pediatr Adolesc Med* 1999;**152**:812–17.

278. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, *et al.* Should unpublished data be included in meta analyses? Current convictions and controversies. *JAMA* 1993;**269**:2749–53.

279. Herxheimer A. Data on file cited in pharmaceutical advertisements: what are they? (The International Congress on Biomedical Peer Review and Global Communications. Prague, 1997.) In: The Cochrane Review Methodology Database; The Cochrane Library (issue 1). Oxford: Update Software, 2000.

280. Davey-Smith G, Egger M. Meta-analysis: unresolved issues and future developments. *BMJ* 1998; **316**:221–5.

281. Clarke M, Oxman AD, editors. Cochrane reviewers' handbook (updated July 1999). In: Review Manager (RevMan), version 4.0 [Computer program]. Oxford: The Cochrane Collaboration, 1999.

282. Duval S, Tweedie R. Trim and fill: a simple funnel plot based method of adjusting for publication bias in meta-analysis. *J Am Stat Assoc.* In press.

283. Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. *Stat Med* 1996;**15**:2493–507.

284. Eberly LE, Casella G. Estimating the number of unseen studies in a meta-analysis [Technical report]. New York: Biometrics Unit, Cornell University, 1997.

285. Klein S, Simes J, Blackburn GL. Total parenteral nutrition and cancer clinical trials. *Cancer* 1986; **58**:1378–86.

286. Kennedy E, Song F, Hunter R, Gilbody S. Risperidone versus "conventional" antipsychotic medication for schizophrenia. In: The Cochrane Library (issue 3). Oxford: Update Software, 1998.

287. Evans S. Statistician's comments on: "Fail safe N" is a useful mathematical measure of the stability of results [Letter]. *BMJ* 1996;**312**:125.

288. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;**50**:1088–101.

289. Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med* 1988;**7**:889–94.

290. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. *Stat Sci* 1988;**3**:109–35.

291. Hedges LV. Modelling publication selection effects in meta-analysis. *Stat Sci* 1992;**7**:246–55.

292. Dear HBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Stat Sci* 1992;**7**:237–45.

293. Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995;**60**:419–35.

294. Cleary RJ, Casella G. An application of Gibbs sampling to estimation in meta-analysis: accounting for publication bias. *J Educ Behav Stat* 1997; **22**:141–54.

295. Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis: a Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Stat Sci* 1997;**12**:221–50.

296. Copas J. What works? Selectivity models and meta-analysis. *J R Stat Soc A* 1998;**162**:95–109.

297. Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. *J Educ Behav Stat* 1996;**21**:299–332.

298. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, *et al.* Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.

299. Pham B, Moher D, Platt R, McAuley L, Jones A, Klassen T. Detecting and adjusting for publication bias in meta-analysis: a systematic review [Abstract]. In: Proceedings of the 7th Cochrane Colloquium; 1999 Oct 5–9; Rome. Rome: Universitás Tommaso D'Aquino, 1999:16.

300. Jones DR. Meta-analysis: weighing the evidence. *Stat Med* 1995;**14**:137–49.

301. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989; **8**:141–51.

302. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140**:290–6.

303. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, *et al.* Methodology and reports of systematic reviews and meta-analyses. A comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;**280**:278–80.

304. Leizorovicz A, Simonneau G, Decousus H, Boissel J. Comparison of efficacy and safety of low molecular weight heparins and unfractionated heparin in initial treatment of deep venous thrombosis: a meta-analysis. *BMJ* 1994;**309**:299–304.

305. Swedish Council on Technology Assessment in Health Care. Cervical cancer (cervix uteri). *Acta Oncol* 1996;**2**(Suppl 7):75–80.

306. Swedish Council on Technology Assessment in Health Care. Uterine cancer. *Acta Oncol* 1996; **2**(Suppl 7):81–5.

307. Swedish Council on Technology Assessment in Health Care. Ovarian cancer. *Acta Oncol* 1996; **2**(Suppl 7):86–92.

308. McAuley LM, Moher D, Tugwell P. The role of grey literature in meta-analysis. (The International Congress on Biomedical Peer Review and Global Communications. Prague, 1997.) In: The Cochrane Review Methodology Database; The Cochrane Library (issue 1). Oxford: Update Software, 2000.

309. Brazier H. Poorly executed and inadequately documented? An analysis of the literature searches on which systematic reviews are based [Abstract]. In: 2nd Symposium on Systematic Reviews: Beyond the basics; 1999 Jan 5–7; Oxford: 38.

310. Bailar JC. Discussion of the paper by Begg and Berlin: Publication bias: a problem in interpreting medical data. *J R Stat Soc A* 1988;**151**:451.

311. Simon R. Discussion of the paper by Begg and Berlin: Publication bias: a problem in interpreting medical data. *J R Stat Soc A* 1988;**151**:459–60.

312. de-Melker HE, Rosendaal FR, Vandenbroucke JP. Is publication bias a medical problem [Letter]? *Lancet* 1993;**342**:621.

313. Moher D. Publication bias [Letter]. *Lancet* 1993;**342**:1116.

314. Scott C, Wasserman T. When is a negative study not negative? *Int J Radiat Oncol Biol Phys* 1997;**39**:859–61.

315. Laupacis A. Methodological studies of systematic reviews: is there publication bias [Letter]? *Arch Intern Med* 1997;**157**:357.

316. Johnston JD, Breimer LH. Publication bias [Letter]. *BMJ* 1991;**337**:1102.

317. Meslin EM. Toward an ethic in dissemination of new knowledge in primary care research. In: Dunn EV, editor. Disseminating research/changing practice (Research methods for primary care, vol. 6). Thousand Oaks, CA: Sage, 1994:32–44.

318. Lynch A. Publication of research: the ethical dimension. *J Dent Res* 1994;**73**:1778–82.

319. Edwards G, Babor TF, Raw M, Stockwell T. Playing fair: science, ethics and scientific journals. *Addiction* 1995;**90**:3–8.

320. Williamson A, White C, editors. The COPE report 1998. Annual report of the Committee on Publication Ethics. London: BMJ Books, 1998.

321. Beveridge WIB. The art of scientific investigation. London: Mercury Books, 1961.

322. Chalmers I. Proposal to outlaw the term "negative trial" [Letter]. *BMJ* 1985;**290**:1002.

323. Gluud C. "Negative trials" are positive! *J Hepatol* 1998;**28**:731–3.

324. Halperin EC. Eureka: it's a negative study. *N C Med J* 1994;**55**:68–9.

325. Petticrew M, Gilbody S, Sheldon TA. Assessing publication bias, and the importance of study quality: the quality adjusted funnel plot [Abstract]. In: 2nd Symposium on Systematic Reviews: Beyond the basics; 1999 Jan 5–7; Oxford: 19.

326. Egger M, Sterne J. Can we detect the presence of publication and related biases? Simulation studies and illustrative examples from the literature [Abstract]. In: 2nd Symposium on Systematic Reviews: Beyond the basics; 1999 Jan 5–7; Oxford: 13.

327. Song F, Rathwell T. Stochastic simulation and sensitivity analysis: estimating future demand for health resources in China. *World Health Stat Q* 1994;**47**:149–56.

328. Fayers PM, Armitage T. Towards an international register of cancer trials: the UKCCCR register of UK trials. *Eur J Cancer* 1993;**29A**:907–12.

329. Zacharski LR, Donati MB, Rickles FR. Registry of clinical trials of antithrombotic drugs in cancer: second report. *Thromb Haemost* 1993;**70**:357–60.

330. Zacharski LR, Donati MB. Registry of clinical trials of antithrombotic drugs in cancer. *Thromb Haemost* 1989;**61**:526–8.

331. Silagy C, Gray S, Fowler G, Lancaster T. Development of a prospective register of smoking cessation trials. *Control Clin Trials* 1996;**17**:235–41.

332. Fayers P, Machin D, Mossman J. Improving communications in clinical oncology: the EuroCODE Project. The EuroCODE Steering Committee. *Br J Cancer* 1992;**66**:607–9.

333. Champney TF. Adjustments for selection: publication bias in quantitative research synthesis. Chicago, IL: University of Chicago, 1983.

334. Patil GP, Taillie C. Probing encountered data, meta-analysis and weighted distribution methods. In: Dodge Y, editor. Statistical data analysis and inferences. Amsterdam: Elsevier Science, 1989:317–45.

335. Rust RT, Lehmann DR, Farley JU. Estimating publication bias in metaanalysis. *J Marketing Res* 1990;**27**:220–6.

336. Larose D, Dey DK. Modeling publication bias using weighted distributions in a Bayesian framework. *Comput Stat Data Anal* 1998;**26**:279–302.

337. Silliman NP. Non-parametric classes of weight functions to model publication bias. *Biometrika* 1997;**84**:909–18.

338. Silliman NP. Hierarchical selection models with applications in meta-analysis. *J Am Stat Assoc* 1997;**92**:926–36.

339. Smith DD, Givens GH, Tweedie RL. Adjustment for publication and quality bias in Bayesian meta-analysis. Fort Collins, CO: Department of Statistics, Colorado State University, 1997.

340. Hedges LV, Olkin LV. Statistical methods for meta-analysis. London: Academic Press, 1985.

341. Hedges LV. Estimating the normal mean and variance under a publication selection model. In: Gleser LJ, editor. Contributions to probability and statistics: essays in honor of Ingram Olkin. New York: Springer Publishing, 1989.

342. Raudenbush SW. Random effects models. In: Cooper H, Hedges LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation, 1994:301–21.

343. Vevea JL, Clements NC, Hedges LV. Assessing the effects of selection bias on validity data for the general aptitude test battery. *J Appl Psychol* 1993;**78**:981–7.

344. General Accounting Office. Cross design synthesis: a new strategy for medical effectiveness research (GAO/PEMD-92-18). Washington, DC: United States GAO, 1992.

345. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technol Assess* 1999;**2**(19).

346. Sutton AJ, Lambert PC, Hellmick M, Abrams KR, Jones DR. Meta-analysis in practice: a critical review of available software. In: Berry DA, Strangle DK, editors. Meta-analysis in medicine and health policy. New York: Marcel Dekker. In press.

347. LaFleur B, Taylor S, Smith DD, Tweedie RL. Bayesian assessment of publication bias in meta-analyses of cervical cancer and oral contraceptives. Fort Collins, CO: Department of Statistics, Colorado State University, 1997.

348. Delgadorodriguez M, Silleroarenas M, Martinmoreno JM, Galvezvargas R. Oral contraceptives and cancer of the cervix uteri – a meta-analysis. *Acta Obstet Gynecol Scand* 1992;**71**:368–76.

349. Moher D, Jadad AR, Nichol G, Penman M, Tugwell T, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;**16**:62–73.

350. Moher D, Cook DT, Jadad AR, Tugwell P, Moher M, Jones A, *et al.* Assessing the quality of reports of randomised controlled trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;**3**(12).

351. Copas JB, Li HG. Inference for non-random samples. *J R Stat Soc B* 1997;**59**:55–9.

352. Taylor S, Tweedie R. A non-parametric "trim and fill" method of assessing publication bias in meta-analysis. Fort Collins, CO: Department of Statistics, Colorado State University, 1998.

353. Taylor S, Tweedie R. Practical estimates of the effect of publication bias in meta-analysis. Fort Collins, CO: Department of Statistics, Colorado State University, 1998.

354. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. An assessment of the impact of publication bias on meta-analyses. Leicester: Department of Epidemiology and Public Health, University of Leicester, 1999.

355. Ashworth S, Osburn H, Callender J, Boyle K. The effects of unrepresented studies on the robustness of validity generalization results. *Personnel Psychol* 1992;**45**:341–61.

356. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;**57**:97–109.

357. Sugita M, Kanamori M, Izuno T, Miyakawa M. Estimating a summarized odds ratio whilst eliminating publication bias in meta-analysis. *Jpn J Clin Oncol* 1992;**22**:354–8.

358. Sugita M, Yamaguchi N, Izuno T, Kanamori M, Kasuga H. Publication probability of a study on odds ratio value circumstantial evidence for publication bias in medical study areas. *Tokai J Exp Clin Med* 1994;**19**:29–37.

359. Spiegelhalter DJ, Thomas A, Best NG, Gilks W. BUGS: Bayesian inference using Gibbs sampling, version 0.5 [Computer program]. Cambridge: MRC Biostatistics Unit, Institute of Public Health, 1996.

Appendix 1

Strategies for searching electronic databases

MEDLINE search strategy

001 exp publication bias/
002 (publication adj2 bias).tw.
003 (selection adj2 bias).tw.
004 (negative adj2 bias).tw.
005 (positive adj2 bias).tw.
006 (submission adj2 bias).tw.
007 (quotation adj2 bias).tw.
008 (retrieval adj2 bias).tw.
009 (reference adj2 bias).tw.
010 (report$ adj2 bias).tw.
011 (unreport$ adj2 bias).tw.
012 (citation adj2 bias).tw.
013 or/1–12
014 (inconclusive adj result$).tw.
015 (null adj result$).tw.
016 (negative adj result$).tw.
017 (publication and bias).tw.
018 16 and 17
019 (nonsignificant adj result$).tw.
020 (non-significant adj result$).tw.
021 14 or 15 or 18 or 19 or 20
022 non-publication.tw.
023 (unpublish$ adj2 trial$).tw.
024 (unpublish$ adj2 data).tw.
025 (unpublish$ adj2 stud$).tw.
026 (unpublish$ adj2 report$).tw.
027 (unpublish$ adj2 paper$).tw.
028 22 or 23 or 24 or 25 or 26 or 27
029 ((report$ adj3 trial$) and publication$).tw.
030 ((report$ adj3 data) and publication$).tw.
031 ((report$ adj3 stud$) and publication$).tw.
032 ((report$ adj3 paper$) and publication$).tw.
033 29 or 30 or 31 or 32
034 (language adj restrict$).tw.
035 ((exclusion adj criteria) and trial$ and bias$).tw.
036 ((inclusion adj criteria) and trial$ and bias).tw.
037 (subject$ adj judg$).tw.
038 (heterogeneity and trial$).tw.
039 ((effect adj size) and bias).tw.
040 34 or 35 or 36 or 37 or 38 or 39
041 21 or 28 or 33 or 40
042 (meta-analysis or review literature).sh.
043 (meta-analy$ or metaanal$ or metanal$).tw.
044 meta-analysis.pt.
045 review academic.pt.
046 review literature.pt.
047 case report.sh.
048 letter.pt.
049 historical article.pt.
050 review of reported cases.pt.
051 review,multicase.pt.
052 42 or 43 or 44 or 45 or 46
053 47 or 48 or 49 or 50 or 51
054 52 not 53
055 animal.sh.
056 human.sh.
057 55 not (55 and 56)
058 54 not 57
059 41 and 58
060 13 or 59

BIDS: Science Citation Index

001 publication bias
002 selection bias
003 negative bias
004 positive bias
005 submission bias
006 quotation bias
007 retrieval bias
008 reference bias
009 reporting bias
010 citation bias
011 file drawer
012 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11

BIDS: Social Science Citation Index

001 publication bias
002 selection bias
003 negative bias
004 positive bias
005 submission bias
006 quotation bias
007 retrieval bias
008 reference bias
009 report* bias
010 citation bias
011 1 or 2 or 3 or 4 or 4 or 6 or 7 or 8 or 9 or 10
012 file drawer
013 meta analysis
014 meta# analysis

015 systematic review*
016 13 or 14 or 15
017 unpublish* trial*
018 unpublish* data
019 unpublish* stud*
020 unpublish* report*
021 unpublish* paper*
022 17 or 18 or 19 or 20 or 21
023 11 or 12 or 16 or 22

SIGLE (grey literature database)

001 selection@bias
002 negative@result#
003 unpublish:@data
004 unpublish:@paper#
005 subject:@judg:
006 unpublish:@stud:
007 bias
008 1 or 2 or 3 or 4 or 5 or 6 or 7

Free text searching all other databases

These included: EMBASE, ERIC, British Education Index, Inspec, Library and Information Science Abstracts, Dissertation Abstracts, Sociofile, Information Science Abstracts and MathSci.

s1 publication bias!
s2 publication(2w)bias
s3 selection(2w)bias
s4 negative(2w)bias
s5 positive(2w)bias
s6 submission(2w)bias
s7 quotation(2w)bias
s8 retrieval(2w)bias
s9 reference(2w)bias
s10 report?(w)bias
s11 unreport?(w)bias
s12 citation(w)bias
s13 s1:s12
s14 inconclusive(w)result?
s15 null(w)result?
s16 negative(w)result?
s17 publication(w)bias
s18 s16 and s17
s19 nonsignificant(w)result?
s20 non-significant(w)result?
s21 s14 or s15 or s18 or s19 or s20
s22 non-publication
s23 unpublish?(2w)trial?
s24 unpublish?(2w)data
s25 unpublish?(2w)stud?
s26 unpublish?(2w)report?
s27 unpublish?(2w)paper?
s28 s22:s27
s29 (report?(3w)trial?) and publication?
s30 (report?(3w)data) and publication?
s31 (report?(3w)stud?) and publication?
s32 (report?(3w)paper?) and publication?
s33 s29:s32
s34 language(w)restrict?
s35 (exclusion(w)criteria) and (trial? and bias)
s36 (inclusion(w)criteria) and (trial? and bias)
s37 subject?(w)judg?
s38 heterogeneity and trial?
s39 (effect(w)size)and bias
s40 s34:s39
s41 s21 or s28 or s33 or s40
s42 (meta-analysis or review literature)/de
s43 meta analysis/de
s44 meta-analy? or metaanaly? or metanal?
s45 dt=meta-analysis
s46 dt=review academic
s47 dt=review literature
s48 case?(2w) report?
s49 dt=letter
s50 dt=historical letter
s51 dt=review of reported cases
s52 dt=review,multicase
s53 s42:s47
s54 s48:s52
s55 s53 not s54
s56 animal
s57 human
s58 s56 not (s56 and s57)
s59 s55 not s58
s60 s41 and s59
s61 s13 or s60

Appendix 2

Data extraction sheet for methodological/ empirical studies

Reviewer _____ Date _____
Author _____
Year _____
Title _____
Source _____

Type

☐ empirical evidence
☐ method/approach
☐ editorial/comment/letter
☐ literature review on P bias
☐ meta-analysis/review
☐ other:

Issues

☐ existence/identifying
☐ magnitude of P bias
☐ consequence of P bias
☐ causes/risk factors
☐ preventing/correcting
☐ other:

Methods

☐ funnel plot
☐ file-drawer method
☐ registration
☐ cohort of registered trials
☐ not applicable
☐ large scale trials
☐ statistical test
☐ modelling/simulation
☐ other:

Areas

☐ general health
☐ other non-health (e.g. education):
☐ specific health (e.g. obesity):

Other relevant information:

Reviewer's commentary (main message, approach, author's conclusion, ...):

Appendix 3

Data extraction sheet for surveying DARE systematic reviews

Acc. Number: _____

Objectives: _____ 1. Effectiveness; 2. Adverse effect; 3. Diagnostic; 4. Other

Type of reviews: _____ 1. Narrative; 2. Meta-analysis

Designs of included studies:

1. RCTs	(Study =	; patients =	)
2. CCTs	(Study =	; patients =	)
3. Other	(Study =	; patients =	)

How were differences between studies investigated?
1. NA; 2. Narrative; 3. Statistical; 4. Meta-regression; 5. Sensitivity/subgroup; 6. Other:

Authors' conclusion:
1. Significant/positive: At least one intervention recommended; or significant difference found between interventions
2. Non-significant/negative: No intervention recommended, or no significant differences found among interventions
3. Unclear: Not able to judge; neither positive nor negative; lack of evidence

Sources searched to identify studies:

1. Not stated	2. MEDLINE	3. EMBASE	4. PsycLIT
5. Cochrane	6. Bibliographies	7. Handsearch	8. Experts/authors
9. Company	10. Proceedings	11. Other:	

Non-English language studies:
1. Unclear

2. Searched	Yes	No	If yes, search methods: _____
3. Identified	Yes	No	How many? _____
4. Included	Yes	No	If included: (a) for main analysis; (b) for sensitivity analysis?

Unpublished studies:
1. Unclear

2. Searched	Yes	No	If yes, search methods: _____
3. Identified	Yes	No	How many? _____
4. Included	Yes	No	If included: (a) for main analysis; (b) for sensitivity analysis?

Issue of publication bias discussed?
1. No 2. Yes

Methods used for dealing with publication bias:

1. Not used	2. Identify unpublished studies
3. Prospective register	4. Fail-safe N
5. Funnel plot	6. Rank correlation
7. Egger's method	8. Large scale trials
9. Modelling	10. Other:

Evidence on publication bias:
1. Not available 2. Available
If available, details: (such as, results of published trials versus unpublished trials; or shape of Funnel plot or related methods)

Any other relevant information:

CCTs, clinical controlled trials; NA, not available

Appendix 4

Stochastic simulation for publication bias

A computer model was established to simulate the consequences of publication bias under different assumptions. The simulated results have been used in chapter 5 to investigate the risk factors for publication bias and in chapter 7 to illustrate the shape of the funnel plots. The model is written in Fortran and consists of the following steps:

Input variables (assumptions):

1. average event rate in the treatment and the control groups
2. number of trials conducted (e.g. 500)
3. the range of the sample size (e.g. 20–1000) with a triangular distribution in which the minimum size equals the most likely size[327]
4. methods for selecting trials for publication (e.g. only studies with $|Z| > 1.96$ being published, or large studies could be selected even if the results are not statistically significant).

Stochastic simulation:

1. randomly determining the sample size
2. randomly determining the event rates in the treatment and the control groups, according to the average event rate and the sample size
3. calculating the lnOR and its SE, according to the simulated event rates and the sample size
4. determining whether the study could be published, according to its statistical significance and/or its sample size
5. repeating the above process for all trials.

Appendix 5

A list of registries of clinical trials

AmFAR AIDS/HIV Treatment Directory[243]

AIDS Clinical Trials Information Service (ACTIS)[243]

AIDS Treatment Resources (ATR)[243]

Anti-platelet Trialists Collaboration (APTC)[243]

ARC Database of Clinical Trials in Rheumatoid Arthritis[243]

California AIDS Clearinghouse (CAC) Registry of Clinical Trials[243]

Cancer Trials Register of the UK Coordinating Committee on Cancer Research (UKCCCR)[243,244,328]

Clinical Trials Registry in Neurosurgery[243]

Clinical Trials Registry of the International Committee on Thrombosis and Haemostasis (ICTH)[243]

Cochrane Controlled Trials Register (CCTR)[213]

DENTALPROJ[243]

Department of Veterans Affairs (VA) Register of Clinical Studies[243]

Early Breast Cancer Trialists Collaborative Group (EBCTCG)[243]

European Computerised Oncology Data Exchange (EuroCODE)[332]

European Organisation for Research and Treatment of Cancer (EORTC) Register of Cancer Trials[243]

International Collaborative Registry of Smoking Cessation Trials[331]

International registers of clinical trials (in: *Cochrane Reviewers' Handbook*)[281]

International Registry of Perinatal Trials (IROPT)[243]

Inventory of Cancer Prevention Trials[243]

metaRegister of Controlled Trials (mRCT)[254]

National Institutes of Health (NIH) Inventory of Clinical Trials[243]

New England HIV Treatment Directory[243]

Ottawa Stroke Trials Registry (OSTR)[248]

Philadelphia AIDS Protocol Testing Directory[243]

Physician's Data Query (PDQ)[243]

Prospective Collaborative Overview of All Current and Planned Randomised Trials of Cholesterol Treatment Regimens[252,253]

Prostate Cancer Trialists Collaborative Group (PCTCG)[243]

Register of Investigative Protocols to treat Malignant Brain Tumours in North America[243]

Register of Randomised Controlled Trials in Primary Care[122]

Registry of Clinical Trials of Antithrombotic Drugs in Cancer[329,330]

Spanish Database of Clinical Trials[243,257]

UK National Research Register[251]

Washington–Baltimore HIV Research Directory[243]

WHO Global Programme for Vaccines and Immunisation Vaccine Trial Registry[250]

Registries of clinical trials under development:

Australian NHMRC Register of Clinical Trials in Cancer and Cardiovascular Diseases[243]

British Heart Foundation (BHF) Register of Cardiovascular Trials[243]

Database of clinical trials of drugs designed to treat serious or life-threatening conditions (USA)[258]

Database of Studies in Complementary Medicine[243]

International Hepato-Biliary Clinical Trials Register[323]

Japanese Foundation for the Multidisciplinary Treatment of Cancer Register of Trials[243]

Registry of Trials in Vision Research[243]

Appendix 6

Modelling for publication bias in meta-analysis

Selection models using weighted distribution theory

Weight functions have been used in various disciplines, including sample surveys and ecology. They are used to adjust results when only partial information is available and the chance of having particular data is related to a feature of the data.[290] Hence, in a meta-analysis setting, weight functions are used to model the selection process and develop estimation procedures that take that selection process into account.[290] There are two aspects to such models: (1) the effect size model, which specifies what the distribution of the effect size estimates would be if there were no selection; and (2) the selection model, which specifies how this effect size distribution is modified by selection.[297] In the models that follow, unless otherwise stated, it is assumed that the chance of a study being included in the meta-analysis is related to the statistical significance of its outcome (implying that journals are more likely to publish significant than non-significant results). In these instances the outcome considered is the observed *p*-value. The appropriateness of such an assumption is explored and alternative approaches, such as conditioning on effect size magnitude, are explored.

General formula for selection model incorporating a weight function

Let T be the observed effect size, and $f(T; \theta)$ the probability density of T irrespective of whether or not the study is published, where θ is the true mean effect size. The weight function is denoted $w(T)$ and the probability density of T, given that the study is published, by $g(T; \theta)$. These expressions are related in equation (1)[205]:

$$g(T; \theta) = \frac{f(T; \theta) w(T)}{A(\theta)} \qquad \text{equation 1}$$

where

$$A(\theta) = \int_{-\infty}^{\infty} f(t; \theta) w(t)\, dt \qquad \text{equation 2}$$

We are interested in the true distribution of T, that is $f(T; \theta)$, and inferences can be made, provided models for $f(T; \theta)$ and $w(T)$ are specified, by constructing a likelihood function and solving it numerically. In the more sophisticated methods, in addition to producing an adjusted treatment estimate, the weight function is estimated through modelling and provides estimates for the selection probabilities that are giving rise to the publication bias. This enables estimation of the probability that a paper will be published, given the *p*-value to be obtained. By examining the magnitude and distribution of the studies' *p*-values, the researcher can assess the chances of selective publication, and hence of bias being present in the meta-analysis. Hedges and Vevea[297] observe that, if the selection model were known, it would be relatively straightforward to obtain an estimate of the unselected distribution of effect size estimates by inverting the selection process. This could be carried out by comparing the observed distribution of effect size estimates to what would be expected and adjusting the selection model to obtain correspondence. Hence, the assumption of the distribution of the effect specified in the effect size model provides the information necessary to estimate the selection model. If the assumption is wrong, the estimated selection model will be wrong, and any corrections based on it may also be in error.

Suggested weight functions

The sections below describe the various selection models proposed for the modelling of publication bias in meta-analysis. They are arranged broadly in order of sophistication, starting with the simplest models. Later sections describe Bayesian methodology to implement weight functions. These methods often use weight functions defined previously from a classic perspective. Although the Bayesian methods are often more difficult to implement, they do have advantages over their classic counterparts, which are discussed below. The first summary section provides an overview of the sections that follow.

Summary of selection models proposed

This section outlines the many different selection models that have been proposed; these are all discussed in more detail below. The simplest is the model of Hedges,[204] which considers the extreme case in which only significant (i.e. $p < 0.05$) studies are observed. This model was implemented in a random effect setting by Champney.[333] Iyengar and Greenhouse[290] provide the first flexible class of selection models in meta-analysis. In these it is assumed that the functional form of the selection model is known. Patil and Taillie[334] extend these models to allow the incorporation of between-study heterogeneity into the model. Rust and colleagues[335] consider weight functions where the effect size, rather than the *p*-value, is considered to affect chances of publication, and a threshold level is estimated from the data rather than being defined *a priori*. Dear and Begg[292] estimate the weight function as a step function and assume that the location of the discontinuities is unknown, while Hedges[291] describes a similar model, except that the location of the steps is assumed to be known *a priori*. These step functions have the advantage that they do not presuppose either monotonicity or knowledge of the direction in which publication bias is operating.[205] Vevea and Hedges[293] expand the previous work of Hedges[291] to allow study level covariates to be included in a mixed model that includes a selection function.

In addition to the classic models summarised above, several authors have used a Bayesian framework. Cleary and Casella[294] consider a simple weight function and apply it to characteristics other than the studies' significance levels. Larose and Dey[336] present a more general method, which can be used to incorporate many of the weight functions described in the classic setting. Although these articles have not developed new selection models *per se*, new developments have been made from the Bayesian perspective. Silliman[337] considers a non-parametric class of weight functions and describes how they can be used to explore the sensitivity of conclusions to its specification. The same author goes on to implement these and other weight functions in a random effects model.[338] This approach allows one to augment the dataset in order to account for unobserved studies. Givens and co-workers[295] consider a similar data augmentation approach and describe a very important application of its use. This approach is then extended to incorporate study quality factors into the analysis.[339]

The simple model of Hedges

The simple model of Hedges,[204] and by Hedges and Olkin,[340] follows on from work conducted by Lane and Dunlap,[203] who carried out a simulation to investigate the degree by which an effect size is overestimated when only statistically significant results are considered. A set of trials varying in size and effect size (measured on a continuous scale) was simulated. To emulate publication bias, only those that were statistically significant were used to estimate an overall effect size. The bias produced by using this procedure was examined as a function of sample size, size of population difference and significance level. As one may expect, the authors concluded that selecting studies in this way can potentially overestimate the effect size considerably. In addition, they observe that the more extreme the α-level required by a journal to publish results, the more distorted are estimates of effect size derived from that journal.

After a simulation parallel to that of Lane and Dunlap,[203] which produced similar findings, Hedges and Olkin modelled the same situation.[340] Studies are dichotomised into those in which significant results (defined by some prespecified level of α) were obtained and hence published, and those that produced non-significant results and were not published. Throughout, these authors take the value of α to be 0.05, and only consider continuous outcomes using the standardised mean difference scale.[340] In some instances (e.g. where relevant studies have been identified and reports retrieved, but effect sizes are not reported if outcome measures are not significant), this model may be realistic; however, more generally, it may be impossible to tell whether or not it realistically models the data. (If a less stringent censoring rule is present, implying that some non-significant results have been obtained, this model can still be used, but the effect size must be calculated using just the results from the significant studies.) Hedges notes that the application of their methods may be misleading if the model is inappropriate.[341]

The weight function takes a value of 1 if the test is significant (at the 5% or any other specified level), and 0 otherwise, expressed mathematically as:

$$w_i(T_i) = \begin{cases} 1 \text{ if } T_i > C_\alpha(v_i) \\ \\ 0 \text{ if } T_i < C_\alpha(v_i) \end{cases} \qquad \text{equation 3}$$

where $C_\alpha(v_i)$ is the critical value of the α-level test for the ith study and v_i is the SE of the ith effect size. (Note that in an effort to keep notation

consistent throughout this section, the original formulae presented by Hedges[204] are not reproduced; instead, the equivalent notation as used by Begg[205] is given.)

A simple solution to this model, based on a vote-counting approach, is initially presented. It treats positive and negative results as independent realisations of a Bernoulli process and the adjusted treatment effect can be estimated using a modification of binomial theory.[204,340] However, this method is limited in its application because it is valid only when a fairly large number of studies exist and when all the studies included are of (approximately) the same size. Owing to these drawbacks, Hedges, with Olkin, considers this approach as most useful for providing quick approximate estimates rather than serving as the analytical tool for the final analysis.[204,340]

A more sophisticated solution is to maximise the likelihood expression formed using equation 1, which can be solved by using computational iteration. Hedges also proposed a simpler method for obtaining solutions to this likelihood function by evaluating it for a grid of trial values and selecting the one that provides the largest value.[204] A further alternative computational method is presented. Rather than simultaneously maximising the likelihood based on all the studies, as described above, one can obtain maximum likelihood estimates for the effect size for each study individually. These can then be combined using a weighted average to obtain a solution. This has the advantages that: (1) no specialist software is needed (a table that provides maximum likelihood estimates for standardised mean differences is provided in the article); (2) the estimates of the treatment effect from each study are calculated explicitly and can be examined directly to detect potential outliers; and (3) this analysis corresponds to the analyses usually carried out in meta-analysis when the results of all studies are observed. It should be noted that no random effect terms are considered in this analysis. In the original article, the method is applied to data from an education example.

A further technical examination of the problem of interpreting the sample mean and variance from a normal distribution when they are reported conditional upon rejection of the hypothesis that the mean is zero is provided by Hedges.[341]

This method has been described as appealing when most of all the published studies are significant, and inappropriate if most of the studies are non-significant.[7] The method is strongly dependent on the nature of the distribution of the *p*-values in the range 0.00–0.05, and accuracy is open to question. The assumption made is that all significant studies are published. Marginally significant ones may not be published; if so, the bias would be underestimated. Begg and Berlin[7] suggest that it would be interesting to investigate study properties further, especially the impact of ignoring the available non-significant publications. Rosenthal comments (in the discussion of Iyengar and Greenhouse's article[290]) that this method assumes the non-published results to have a mean effect of 0, which is probably too simple. There is evidence that suggests this mean is displaced in the direction of the mean of the published studies; the maximum likelihood estimate approach of Iyengar and Greenhouse[290] addresses this issue by trying to estimate the mean effect size in the population (see below). Clearly, in the instance where experiments are identified but the available study reports do not give effect sizes owing to lack of significance (Hedges refers to this as reporting bias) this model may be useful. However, in health technology assessment and related disciplines, this scenario is not usual, the major concern being with studies for which no account has been published. In such a situation a more realistic model is required.

Champney's weight functions

Champney[333] considered the same situation as Hedges,[204] where it is assumed that all studies reporting significant results are published and all studies with non-significant results are not. Champney extends Hedges' work by developing a method of adjustment, which is based on the random effects model. Maximum likelihood estimates are derived by using both grid-searching techniques and the expectation–maximisation algorithm. The model used assumes that the random effects are normally distributed. The method developed is investigated via Monte Carlo simulation studies and applied to a psychotherapy meta-analysis dataset.

This work suggests that publication bias may have substantial effects on estimation of the between-study variance, even when the estimate of the mean is not strongly affected (see Hedges' comment on the article by Iyengar and Greenhouse[290]).

Weight functions of Iyengar and Greenhouse

Iyengar and Greenhouse[290] expand upon the model of Hedges by presenting more sophisticated censoring schemes. Two different families of weight functions (w) are discussed. These are described algebraically below:

equation 4

$$w(x;\ \beta,\ q) = \begin{cases} \dfrac{|x|^{\beta}}{t(q,\ 0.05)^{\beta}} & \text{if} \quad |x| \le t\,(q,\ 0.05) \\ 1 & \text{otherwise,} \end{cases}$$

equation 5

$$w(x;\ \gamma,\ q) = \begin{cases} e^{-\gamma} & \text{if} \quad |x| \le t\,(q,\ 0.05) \\ 1 & \text{otherwise.} \end{cases}$$

In both cases, $p\big(|T_0| \ge t\,(q,\ 0.05)\big) = 0.05$, where T_0 has a central t distribution with q degrees of freedom. Both of these functions (equations 4 and 5) imply that all studies that are statistically significant at the 0.05 level will be published because the weight functions will take the value 1 over these values. When β and γ are zero, the weight functions indicate no selection bias and they approach the weight function of Hedges described above as they approach infinity.[204] For non-significant results, equation 5 considers the reporting probability as constant, but not zero (as was the case for Hedges' model). Equation 4 implies that the reporting probability increases as the outcome approaches statistical significance. Varying the values of β and γ changes the "severity" of the selection model (see original article for details).

The likelihood created when these weight functions are combined with the density function in equation 1 can be solved by using maximum likelihood methods. β and γ are estimated through the modelling and do not need to be specified beforehand.[290]

The authors comment that this method is flexible, and that one can perform sensitivity analyses by varying the assumptions and examining the log likelihood surface, which shows how informative the data are about the parameters in the model. Iyengar and Greenhouse's original article[290] was published with an extensive commentary by several authors. This commentary is summarised below.

Hedges agreed that their model was more realistic than his own previous attempt and the work of Champney[333] (see above), although he has commented that his model still has a role in providing a simple method for producing an upper bound on the biasing effects of selection.[341] However, he still considers them as unrealistic because other factors apart from the p-value, such as size and study design, also play an important role in the decision to publish, from both researchers' and journal editors' perspectives. Hedges argues that well-designed studies with no effect will generally be published, but small poorly designed ones with very small p-values may not. It is when the p-value has an intermediate value that it may greatly influence the decision to publish. For this reason he suggests that an s-shaped curve similar to a logistic function may be better, yielding:

$$w(x;\ \alpha,\ \beta) = \frac{e^{\alpha + \beta|\theta|}}{(1 + e^{\alpha + \beta|\theta|})} \qquad \text{equation 6}$$

where θ is the effect magnitude, α sets the probability that a result is observed even when $\theta = 0$, and β is a constant that determines the slope of the curve near its inflection point.

Hedges also suggests that it would be interesting to investigate the effect on estimates in random effects models using the more sophisticated weight functions of Iyengar and Greenhouse. He comments that there is little empirical evidence to guide the choice of weight functions, but one approach may be to obtain estimates and p-values from unpublished studies located in study registries. He goes on to conclude that selection models should be used as a form of sensitivity analysis; estimates from a variety of selection models are valuable to assess the sensitivity to selection effects on the conclusions of the meta-analysis.

Rosenthal and Rubin comment that the notion that only significant results found in one direction are published is probably unrealistic and argue that a two-tailed selection process (which is adopted in Iyengar and Greenhouse's selection models) is more realistic. The reasons given for this are: (1) early in the history of a research domain, results in either direction are important news; and (2) later in the history of the domain, when the preponderance of the evidence has supported one direction, significant reversals are often more important news than are further replications.

Along similar lines, McPherson (discussion of Begg and Berlin's article[7]) comments that she would like to impose asymmetry on the publication criteria. Significant results in the expected direction will have a different impact from significant results in the opposite direction (for example, beneficial effects of new therapies are more likely to be published than ones with deleterious effects).

Laird and colleagues commented that symmetry in these models, and uniform weight in the tails, may not be completely realistic. They provide some results to substantiate their claims. Additionally, these authors provide formulae for estimating the number of unpublished studies by using weight functions.

Bayarri discusses the potential of using Bayesian methods to take into account both the effect that selection has on the model and the influence that the number of unobservable, unpublished studies might have in the final conclusion of the analysis.

In the rejoinder, Iyengar and Greenhouse respond to the criticism that they used a fixed and not a random effect model, by commenting that they do advocate a model that takes into account between-study heterogeneity. However, in order to keep the modelling simple, they presented selection model methodology only from a fixed effect perspective. In addition, the authors present a third weight function, which addresses concerns about whether the lack of sensitivity to the choice of weight function about $t = 0$ and to the fact that both weight functions give uniform weight to the tails. This asymmetrical function is described below:

$$w(x;\ \alpha, \beta) = \begin{cases} 1 \text{ for } x > t(q, 0.05), \\ e^{-\alpha} \text{ for } |x| \leq t(q, 0.05), \\ e^{-\beta} \text{ for } x \leq -t(q, 0.05) \end{cases} \quad \text{equation 6a}$$

where $t(q, 0.05)$ is defined in equations 4 and 5 above.

The method is illustrated briefly using data from a meta-analysis investigating the effects of open versus traditional education on creativity, and results are compared with those previously obtained by Hedges and Olkin by using the Hedges model.[340]

Weight functions of Patil and Taillie

Patil and Taillie[334] extend the selection models described by Iyengar and Greenhouse[290] by implementing their weight functions in a model that is capable of accommodating between-study heterogeneity. Their method is illustrated with a dataset previously considered by Hedges and Olkin[340] and Iyengar and Greenhouse[290] on the effects of open versus traditional education on creativity. In this analysis, additional weight functions are examined; the feasibility of using a normal approximation to the non-central t-distribution is explored; and heterogeneity is incorporated into the analysis. The authors observe that heterogeneity and publication bias are two competing explanations for the overdispersion of test statistics. The analysis concludes that, in this example, the normal approximation is quite adequate, and that heterogeneity accounts for nearly all the overdispersion with virtually no evidence of publication bias being present. The authors comment that further investigation is required to determine the general suitability of the normal approximation and to determine if such decisive discrimination between heterogeneity and publication bias is always possible.

Weight function of Rust, Lehmann and Farley

Rust and colleagues[335] describe a weight function for estimating publication bias based on a measure of effect size, rather than on the significance of the associated p-value, as described previously. This article has largely been ignored by other researchers in the field, perhaps owing to the economic context in which it was published, although the model developed is generalisable to the treatment estimates found in medicine and its related disciplines. The model assumes that publication bias involves a fixed censorship threshold, beyond which no censorship occurs. This censorship threshold is estimated in the model rather than specified *a priori*, which implies that, if the reported effect size is greater than the threshold value, the probability of publication is 1 and, when it is smaller than this threshold, the chances of publication are given a fixed probability value, which is also estimated from the model. A likelihood ratio test derived from a constrained version of the general censorship model is described. The authors state that this model can test different underlying assumptions about the parametric form of the underlying density. Model parameters are estimated via maximum likelihood methods.

The performance of two differently shaped distributions, namely the exponential and the Erlang 2 distribution (rather than the normal distribution typically used) for the distribution of treatment effects (i.e. for $f(T; \theta)$ in equation 1) are explored via a simulation study. This simulation concludes that the proposed method is conservative, tending to err more on the side of not detecting censorship, but generally produces sensible results. Two comp uter programs have been developed to fit the model and are available on request. The application of the model is illustrated by using two previously reported meta-analyses. Although the model seems to work very well, the authors do note some limitations. One

of these is the fact that the method does not adjust for differences in the sample size across the studies, although they state that this would be relatively easy to incorporate.

Hedges' stepped weight function

Hedges generalised the weight function method of Iyengar and Greenhouse (see above).[291] Instead of specifying a parametric form for the weight function, a step function is used, which allows the weight function to take different values in different regions of the *p*-value scale. The value on each interval of the step function is then estimated. Unlike Iyengar and Greenhouse, who use a fixed effect model, Hedges here adopts a random effect meta-analysis approach.

The evidence for using a step function, and deciding on the placement of the discontinuities came from information provided by psychological studies related to researchers' perceptions of *p*-values and conclusions drawn. The reasoning used is that a study with a *p*-value of 0.01 is more likely to be published than a study with a *p*-value of 0.05, which, in turn, is more likely to be published than a study with a *p*-value of 0.10 etc. These "cliff effect" milestones in significance are reasoned as realistic because a result with a *p*-value of 0.045 is perceived as much more conclusive than one with a *p*-value of 0.055; however, a pair of results with *p*-values of 0.045 and 0.035 are perceived as about equally conclusive.

The weight function is modelled in such a way that it permits the data to reveal the shape of the function; however, the steps are determined *a priori*. Plotting *p*-values may provide insight into the probable shape of the weight function, by suggesting where the discontinuities lie. This may also provide insight into the likelihood that publication bias exists. The general form of the weight function, for a two-sided test (which assumes studies with significant *p*-values in either direction are more likely to be published; if one were examining a funnel plot, this would correspond to a deficit of studies in the centre of the plot) is given by:

equation 7

$$w_i(T_i) = \begin{cases} w_1 \text{ if } 0 < |T_i| < C_{a(1)}(v_i) \\ w_2 \text{ if } C_{a(1)}(v_i) < |T_i| < C_{a(2)}(v_i) \\ . \\ . \\ w_i \text{ if } C_{a(l-1)}(v_i) < |T_i| < \infty \end{cases}$$

where the notation is defined for equation 3 and critical values $\alpha(1)\ldots\alpha(l-1)$ are the critical values selected by the researcher.

The parameters can be estimated via maximum likelihood methods. The formulation in the original paper[291] employs two-tailed *p*-values, while later formulations[293,297] have employed cutpoints determined by one-tailed *p*-values.

As well as this estimation procedure, tests for publication bias, which assess if all the weights are equal to 1, are presented. The first is based on the chi-squared statistic, and the second on a likelihood ratio test (see original article for details). The author stresses that the tests are conditional in that they depend on the assumption that the random effects are normally distributed and that the true weight function can be adequately approximated by a step function with steps that were specified *a priori*.

Despite its being substantially more sophisticated than previous models, Hedges still appears cautious about the method, suggesting it should be used to give a "broad indication of whether selection is operating".[291] It should be noted that at least one observed *p*-value needs to exist for every step interval defined.[292]

Later, this method was modified in an article by Hedges and Vevea.[297] Here, the model is altered to reflect a selection process based on one-tailed *p*-values, instead of the two-tailed *p*-values described above (see original article for model specification).

This article also contains an extensive investigation into the applicability of such models in meta-analysis. Simulations were carried out to examine how well the procedures perform when the distributional assumption of normal random effects is met, and how robust they are to violations of that assumption. A *t*-distribution with 3 degrees of freedom was used to examine the procedure's behaviour when the distribution of random effects has much heavier tails than the (normal) model assumes. A platykutyic distribution was also considered; this comprises a mixture of two normal distributions with means separated by two standard deviations. This was done to mimic a meta-analysis in which the outcomes represent a mixture of results from two similar but distinct populations, and the researcher fails to detect the mixture. In addition, further mixture and skewed distributions were investigated. These distributions of effect sizes were studied by using a weight step function with four intervals, and four selection conditions were examined, ranging from the null model, where no selection occurred, through light, and moderate, to an extreme selection scheme. The bias of the maximum likelihood estimators

and the adequacy of 95% CIs based on them were calculated for an array of simulations based on the distributions of effect size, weight function and degree of selection specified above. The mean squared errors of the estimates were also calculated because the use of the selection model increases variability in the pooled results.

The authors conclude that the procedure does a "remarkably good job" of estimating the selection model, even when the model for the random effects is (modestly) mis-specified. However when the SEs of the treatment effect estimates from individual studies are large, the estimated selection model may be too uncertain to be useful for some purposes. This is because the data contain relatively little information about the weights in the selection model. In addition, the procedure is also effective in reducing bias in the mean effect, even when the model for the random effects is modestly mis-specified. However, there does appear to be a tendency to underestimate the uncertainty of the mean effect size when selection is present (i.e. CIs are too narrow). This leads the authors to suggest that unless the selection, and hence the bias, are judged to be profound, the uncorrected estimates may be more accurate.

The authors comment that the most surprising aspect of these results is that the procedure performed so well when the random effects were mis-specified; hence this robustness to model mis-specification makes this model useful. However, the likelihood ratio-based significance test for selection performed only modestly when the distribution of the random effects was correctly specified, but was "abysmally" non-robust.

Application of Hedges' model

Hedges presented an example of the use of the method where it is applied to data from a meta-analysis of 755 studies of the validity of the General Aptitude Test Battery, a cognitive test measuring several mental abilities.[291] The data reported for each scale consisted of a correlation coefficient between a General Aptitude Test Battery ability scale and a measure of job performance. Analysis was performed on the Fisher z-transformed correlations. In the interval $0.5 \leq p \leq 1.0$, only 70% of the expected number of p-values were observed; however, the overall pattern of p-values did not differ significantly from that expected if there was no publication bias. Hence, the pooled estimate adjusted for publication bias did not differ significantly from the standard random effect result.

Semi-parametric weight functions of Dear and Begg

Dear and Begg [292] suggested that the problem with the above weight functions (excluding Hedges' step function) is their lack of flexibility for accommodating different shapes of selection functions; a particular problem is the constraint of monotonicity. They presented an approach that allows the shape of the weight function to vary in as unconstrained a manner as possible, using a semiparametric model. This model also has the ability to incorporate random effects for outcome. The main distinction between this approach and the step function of Hedges (outlined above) is that Hedges chooses to prespecify the regions of the p-value scale within which the weight function is assumed to be constant (cutpoints of 0.05, 0.01, 0.001 etc.). Here, the steps in the weight function are data driven. A two-sided p-value approach is taken throughout the article; however, the authors note that the methodology can be adapted easily to the one-sided setting. The general form of the weight function is given by:

equation 8

$$w_i(T_i) = \begin{cases} w_i \text{ if } & \infty > |T_i| \geq -v_i\Phi^{-1}(p_{2l-2}/2) \\ . & \\ . & \\ w_j \text{ if } & -v_i\Phi^{-1}(p_{2j}/2) > |T_i| \geq \\ & -v_i\Phi^{-1}(p_{2j-2}/2) \\ . & \\ . & \\ w_l \text{ if } & -v_i\Phi^{-1}(p_{2l}/2) > |T_i| \geq 0 \end{cases}$$

where $\Phi(\cdot)$ is the standard normal distribution function. The p-values $p_1,...,p_{2l}$ are ranked and grouped in twos because the model would be non-identifiable if the weight function were allowed to vary between adjacent, ranked p-values. Parameters can be estimated by using maximum likelihood methods. It is important to note that only the relative weight function, and not the absolute weight function, is calculated in this method. (The latter would require information on the p-values of the unpublished studies). Hence, the weight function produced is used to provide a visual display of the relative weight function for the purposes of identifying publication bias, and is not used to adjust the pooled estimate. Like the step model of Hedges described above, the estimated relative weights can be used in a formal test for bias using either a rank correlation or a likelihood ratio test.

This method is complementary to the traditional funnel graph but, unlike the funnel plot, it is sensitive to publication bias even when the study

sample sizes are similar. The authors stress that this model should be used as an exploratory, informal tool and suggest that it can be used to correct estimates for bias, or preferably to focus attention on the causes of bias.

Dear and Begg note that this method could be adapted to other metric scales. For example, if it were felt that the observed effect, rather than the *p*-value, was the primary determinant of bias, this could be used instead. This also holds true for the other weight functions considered above. The authors consider how the results of their method differ from Hedges' step function. They say Hedges' method will lead typically to a weight function with fewer "steps" and, as a result, is probably more robust but less flexible than their own method. They suggest that further research is required, but intuitively note that Hedges' model will be more suitable for meta-analyses with substantial numbers of component studies, while their method will be necessary for small meta-analyses.

General linear model using weight functions of Vevea and Hedges

Vevea and Hedges[293] describe general linear models, which can include fixed and random effect terms, where selection is based on one-tailed *p*-values (these are sometimes referred to as mixed models[342]). Essentially, this extends the previous work of Hedges[291] to allow the inclusion of linear predictors in the model for unselected effects. This permits the distinction between systematic variability from general heterogeneity, and avoids the possibility of heterogeneity being confused with selection bias. The same selection model is used as by Hedges,[291] although a one-sided *p*-value approach is taken here.

Maximum likelihood techniques are used to solve the model. In addition to providing adjusted estimates for the model parameters, this method enables one to construct likelihood ratio tests for differences in fit among different specifications with different constrained parameters. It is also possible to test a selection model with a model assuming no selection effect to assess whether, in fact, publication bias is present. With a generalisation of the method, it would be possible to select different weight functions for various types of studies included in the meta-analysis. A computer program for implementing this methodology is available from the first author (Vevea).

Applications of Vevea and Hedges' method

The methodology was applied to a large meta-analysis dataset (489 studies)[293] examining the efficacy of psychotherapy in the treatment of phobia. The analysis uses a standardised mean difference outcome scale. Two binary covariates were included in the model to indicate whether the treatment was behavioural or due to desensitisation and if the phobia was complex or simple. The interaction between these two covariates was also considered. After this model was fitted and estimates obtained, the selection model, defined as a 10-level step function with discontinuities at milestone *p*-values, was applied. When this model was fitted to the data, the estimated height of the steps seemed reasonable because, in general, they decreased as the significance band defined decreased, suggesting a higher degree of selectivity for the least significant studies. The likelihood ratio test for the addition of the weights to the model was significant, indicating that the weight function substantially improved the fit of the model. Hence, adding a selection model to the generalised linear model produced updated coefficients for the covariates as well as the overall pooled estimates. The estimated effects for the four possible treatment/phobia conditions (defined by the covariates) reduced by 15–25%, while the SEs of these predicted means were two to three times as large as they were before the weight function was estimated.

This method has also been used in a meta-analysis of placebo-controlled homoeopathy trials.[298] This is a rare example of a selection model being used to adjust the result in a reported meta-analysis published in a medical or related journal. The funnel plot of the 89 studies being meta-analysed suggested that publication bias may be present (the left hand tail being smaller than the right, suggesting the potential omission of small non-significant trials). Egger and co-workers' test[44] was significant, which confirmed that publication bias was present and, as a result, the treatment effect would be artificially inflated. If the file drawer method[5] was used, this assumed that all missing RCTs had an average odds ratio of 1 (i.e. these trials showed null results on average), and it was found that 923 missing trials of average size (118 patients) would be required to reduce the pooled effect size to insignificance at the 0.05 level with the random effects model. When a selection model of the form described above was applied, the pooled OR changed from 2.45 (95% CI 2.05 to 2.93) to 1.78 (95% CI 1.03 to 3.10), a decrease of 27%. However, after this adjustment for publication bias, the result remained highly significant and, because a very large number of unpublished non-significant studies would be needed to overturn this result, it would seem

that the treatment benefit was not entirely due to publication bias. As an aside, the authors comment that in this instance the interpretation of the funnel plot is difficult because they would not expect the effect of homeopathy to be homogeneous for all clinical conditions, and hence the plot in effect comprises multiple overlapping funnel plots (some centred around no effect, others around real non-zero effects). The utilisation of three methods to assess publication bias is also noteworthy in this example.

Another meta-analysis has been published, which used this method to explore the impact of publication bias.[343] This meta-analysis included 755 studies, which investigated the validity of the General Aptitude Test Battery by measuring its correlation with criteria relevant to job performance. Previous analyses had suspected that publication bias was present, and it is known that all studies carried out were not included. Analysis suggested that a moderate level of publication bias was evident, but this would have only a small effect on the majority of the pooled subscales (reducing the correlation by approximately 10%); however, the correlation in the Clerical Perception subscale reduced the correlation by a more substantial 30%. Hence, in general, the conclusions of the meta-analysis did not change after adjusting for publication bias.

Weight functions implemented from a Bayesian perspective

The sections below describe the use of selection models from a Bayesian modelling perspective. This is an alternative to the often-employed classic formulation. Bayesian methods have gained in popularity over recent years, partly due to their appealing nature and partly as a result of advances in computational methods, which make possible their practical implementation.

Methodology of Cleary and Casella

Cleary and Casella[294] consider the simple selection model approach from a Bayesian perspective. The weight function used is equivalent to equation 4, which assumes that all studies significant at the 5% level are published, while all non-significant results have a constant probability of being published. Gibbs sampling was used to produce posterior estimates for the parameters of interest.

The authors go on to consider a model in which the size of the selection parameter for an individual study is determined by some of the characteristics of the study itself (i.e. where the probability of a non-significant study being published is based on study characteristics). An approach, difficult to implement in practice, is to assign a different prior distribution for the probability of publishing for every study. However, the approach taken focuses on what characteristics of the study make one consider that it was more or less likely to be published, and then to build a model that explicitly includes those characteristics. Possible characteristics include sample size, amount of funding, reputation of the journal, or previous work of the authors. It should be stressed that no method is given for considering more than one factor simultaneously (e.g. *p*-value and sample size). The development of suitable joint priors for study characteristics that are not independent might be difficult, except through maximum likelihood estimation.

Application of the methodology

The method developed is illustrated with an example from the field of education research, specifically the value of coaching for students preparing to take a Scholastic Aptitude Test. When the analysis is adjusted for publication bias, either through *p*-values or sample size, the significance of the pooled result diminishes. It is interesting that this example includes only four studies, none of which is significant at the 5% level, which makes it an unusual choice of example to illustrate the method.

Larose and Dey weight functions

Larose and Dey[336] provide a unified methodology for a Bayesian view of modelling publication bias. They investigate a range of statistical models describing possible mechanisms of publication bias, which incorporate weight functions in a natural way. The sensitivity of the overall effect estimate (as well as estimates of the other model parameters) can be studied over a range of models. The authors suggest it is preferable to demonstrate the robustness of these effect estimates over several models, rather than to depend on a single model. A Bayesian framework is presented; within it the weight functions of Iyengar and Greenhouse,[290] Patil and Taillie,[334] and Hedges[204] are considered. The authors comment on the appeal of being able to produce posterior distributions on which inferences can be based by using a Bayesian approach. Non-informative priors are used throughout analyses. Markov Chain Monte Carlo integration methods are used to implement the models, and details are provided in the original article.

Five candidate weight function are considered, including a constant weight function that ignores the possibility of publication bias. Bayesian model selection criteria are then applied (using conditional predictive ordinate plots), in order to determine which weight function is most favoured by the data.

Application of the method

The education data example first described by Hedges[204] and Hedges and Olkin[340] is used to illustrate this method. Several questions are addressed in the analysis. First, concerning the presence of publication bias, it is possible to look at the 95% credible intervals for a parameter in the weight functions to ascertain whether it includes zero or not. If it does not, there is evidence of publication bias. The variation in the overall pooled estimate over the different models may be examined to indicate the robustness of the result and the impact of publication bias. It is also possible to go one stage further and take a weighted average of the results produced by the various selection models. Three different weighting schemes for doing this are discussed. An investigation into which model fits the data best is carried out using conditional predictive ordinate plots (see article for details). Finally, these results and those from classically-based maximum likelihood estimates are compared; however, owing to the fixed effect nature of the classic methods, the results are not from equivalent models and hence are not directly comparable. They conclude that the Bayesian formulation is preferable because of the array of questions it addresses (discussed above).

Non-parametric weight functions of Silliman

Silliman describes a non-parametric ε-contamination class of weight function.[337] The illustration of how this class of weight function can be used to explore the sensitivity of conclusions to its specification is the main aim of the report. This takes weight function specification one step further than previously, when the weight functions used were either completely specified or had partly specified (semiparametric) forms. Here, non-parametric functions are explored. The author claims that this provides a robust method for specifying publication bias in a meta-analysis and determining whether conclusions are sensitive to this bias. The method involves consideration of a weight function of general form:

equation 9

$$w(x;\ \varepsilon,\ q,\ w_0) = \{w : w = (1-\varepsilon)w_0 + \varepsilon q,\ q \in Q\}$$

where w_0 is a base weight function specified by the investigator, and different specifications of Q produce different ε-contamination classes; ε can be thought of as the degree to which one has a lack of confidence in the choice of the base weight function. Hence, one can see how the quantities of interest change as the weight function, w, varies over the entire class $w(x;\varepsilon,q,w_0)$. The only restriction placed upon the weight functions considered is that they are non-negative. To maintain interpretability of the weight functions, only classes that are uniquely defined are considered, and a specification method to avoid non-uniqueness is described in the original report. A Bayesian approach is taken, hence a prior distribution is specified for the unknown parameter(s) in the weight function and the posterior distribution is computed. The use of sensitivity analysis is important when combining information; two reviews of the literature have appealed explicitly for its increased use.[344,345] This method provides a tool for investigating the robustness of results to the specification of the weight function in meta-analysis. It should be noted that the method is computationally demanding and Silliman recommends that it should be limited to simple classes of model. Disappointingly, no application of the method to an existing meta-analysis dataset is yet available.

Bayesian hierarchical selection models and data augmentation as described by Silliman

In a further report to the one described above, Silliman discusses weight functions in the context of hierarchical selection models for meta-analysis.[338] This approach combines the use of hierarchical models (often called random effects models), which allow one to investigate variability both within and between studies, and also weight functions, as described above. Known, semiparametric and non-parametric weight functions are all considered, and special attention is given to models related to that of Hedges.[291] This method is extended to combine the selection model with data augmentation approaches (see below) in order to account for unobserved studies. This latter method is used to assess the sensitivity of results to any unobserved study effects. A Bayesian approach is used throughout, and estimation is carried out via Markov Chain Monte Carlo integration. A framework for including weight functions into a Bayesian random effects meta-analysis is described but, owing to its length and complexity, it is not reproduced here. The author observes one advantage this method has over the classically-based method of Hedges[291] is

that here posterior distributions for the parameters of interest are produced.

This data augmentation procedure extends the work of Rosenthal[5] and Iyengar and Greenhouse,[290] and attempts to estimate the number of unobserved studies. Using the hierarchical selection model approach together with data augmentation accounts simultaneously for heterogeneous study effects and bias involved in the collection of studies, and also addresses the sensitivity of the final conclusions to any unobserved studies. Additionally, unlike either of the previous approaches referenced above, unobserved studies are no longer restricted to being non-significant at some chosen significance level. The "augmented" studies are those that were generated and then not published/identified/selected by the researchers carrying out the meta-analysis according to the selection mechanism defined by the weight function.

The conclusion to the original report states that the implementation of the methods described is straightforward. However, we suggest that the methods used are complex and, as for many of the methods described in this section and as noted previously,[346] no standard software exists for implementing them. Researchers other than expert statisticians are likely to experience problems in executing this method in practice.

Illustrative example

An example from dentistry is provided to illustrate the method. Twelve studies that compare the effectiveness of two different types of fluoride in preventing tooth cavities are investigated. The outcome considered is the average increment in decayed, missing and/or filled surfaces of teeth; it is measured on a continuous scale. This example provides a very clear exposition of the Bayesian approach to meta-analysis generally, in addition to providing insight into the selection model and data augmentation extensions described. Consideration is given to a realistic prior distribution for parameters in a number of parametric selection models, which are examined over a range of beliefs about how extreme was the selection bias.

The data augmentation approach of Givens, Smith and Tweedie

Givens and colleagues[295] discuss a Bayesian data augmentation approach to adjusting the analysis for publication bias. Weight functions related to those of Hedges[204] are used in this model. There are similarities between this method and that of Silliman[338] described above. The approach augments the observed data by simulating the outcomes for the missing studies, creating a "complete" dataset for analysis. Gibbs sampling is used to obtain posterior densities for the parameters of interest. The technical details of implementing the model are lengthy and complex (and so are omitted here). No software is currently available to implement this method, which requires a moderate number of studies to be included in a meta-analysis.

The original article also describes simulation studies to evaluate the reasonableness of the method; these conclude that the method is somewhat sensitive to the choice of prior placed on the probability of publication in each interval of the step weight function used. However, the impact on the final outcome estimate (RR in this case) is less serious than the impact on the number of imputed studies might indicate. The authors comment that using a semiparametric weight function should be more robust than a parametric method to changes in the form of the exclusion criteria.

Application of the method

This approach is illustrated with an important epidemiological example concerning the relationship between exposure to environmental tobacco smoke and lung cancer. The US Environmental Protection Agency published a report that included a meta-analysis of 31 studies on the association of lung cancer in people who had never smoked with environmental tobacco smoke exposure through spousal smoking, suggesting there was a link between the two. Several tobacco companies filed a lawsuit against the Environmental Protection Agency claiming that various sources of bias, including publication bias, could explain the association claimed. The original analysis did not make any adjustment for publication bias. In this article the methodology developed is used to assess the impact of publication bias on the result of the meta-analysis.

When applied to world-wide data, the analysis indicated that there may be around ten possible missing negative studies and a similar number of missing insignificant positive studies. After allowing for these, the 95% posterior credibility interval for the RR is shifted downward towards the null hypothesis of no effect, and the actual estimate of excess risk is cut to approximately one-third of the unadjusted value.

A written discussion of this article has been published.[295] In this, Begg makes the comment

that many of the epidemiological studies included in the environmental tobacco smoke example were case-control studies and there is a distinct possibility that a proportion of these were driven by the data opposed to the studies' primary hypotheses. By this he is suggesting that data on exposure to many different risk factors could have been collected in studies and, after all these were examined, only the statistically significant ones were chosen for publication. This adds weight to the argument that publication bias could exist, and suggests that it could be a greater problem in epidemiology than meta-analyses of RCTs. Begg does not generally support the use of methods that adjust via selection models, commenting:

> "My own experiences with this kind of approach lead me to believe that it is not a sound basis for making inferences about the true effect size, and that these models are useful only as part of a set of semiformal tools for identifying bias rather than for correcting it."

He also considers that the approach taken here is overcomplicated and does not add insight to the analysis. A general technical issue raised by Begg is that random effects meta-analyses (used here) are much more susceptible to publication bias than the fixed effect approach. This is because the random effects model gives more weight to the smaller studies in the analysis, which are more likely not to have been published, and hence the estimate produced will have greater bias in the presence of selective publication. In addition, because publication bias tends to exaggerate the apparent heterogeneity, this further accentuates the bias in the random effects estimator.

DuMouchel and Harris[295] are more supportive of the developed method but acknowledge that it is difficult to comprehend the technical details. They go on to examine the model used in detail and question why selection was based on one-sided (rather than two-sided) p-values. In addition they comment that the dependence of the selection criterion should probably be based on more than the p-value. They suggest that the sample size, cost or power of a study may be additional selection criteria. Specifically, they suggest adding a factor for the dependence on the SE of each study's estimate to the selection model to assess the impact this would have on the results.

Dobson and Dear[295] are supportive of the method, stating that: "Now that meta analysis is taking a high-profile role in public policy-making and regulatory affairs, it is entirely appropriate that more sophisticated techniques, such as [these]... be developed." They consider technical details about the analysis and discuss possible choices for priors for the weights in the selection model. A further suggestion is that, if publication depends on sample size as well as p-value, then the augmented dataset should reflect these patterns and should include studies of different sizes with different frequencies.

In the rejoinder that follows these comments, the original authors were generally supportive of the suggested extensions to the methodology.

Extensions to the data augmentation approach of Givens, Smith and Tweedie

In a further report, Smith and colleagues[339] extend the above methods to incorporate study quality effects into the adjustment for publication bias. This method aims to remove the limitation that selection is modelled solely as a function of significance level or effect size, and also incorporates a quality factor into the model. Studies are divided into several tiers; in the report these are based on quality classifications of the studies, but they could refer to any other study characteristic. In addition, these groups could be further subdivided to allow consideration of multiple factors.

The model implemented is compatible with the belief that statistically significant results are more likely to be published than non-significant studies; but it also assumes that the selection might be based on the tier in which the study falls, so that, for example, a non-significant high-quality study may be accepted with a higher probability than a non-significant low-quality study. As before, weight functions of the form described by Hedges[204] are used. Simulations to investigate the performance of the method are described in the report. These conclude that the method performs well when publication bias is present, but not as well as a standard Bayesian analysis when it is not.

Application of the extensions

This method is applied to a meta-analysis investigating the association between cervical cancer and oral contraceptive use.[339,347] Many studies have been carried out on this topic. The results range from very strong associations to negative associations. A previous meta-analysis[348] identified 62 published relative risks from 51 published articles. These results were classified into two groups: group 1 – all 62 results, and group 2 – 26 'methodologically acceptable' results of higher quality studies. When the original method[295] is applied to group 2, there appears to be publication bias present; imputation reduces the estimated risk while narrowing the posterior CI.

The data were then analysed using the multi-tier approach. In doing this it is assumed that there is sufficient information in the low-quality studies that they should not be omitted lightly from a meta-analysis, but that it may be reasonable to assume that the low-quality studies have a different chance of being published compared with the high-quality group 2 studies. Grouping the studies into quality tiers allows adjustment for publication bias within each tier, with the hierarchical structure of the model reflecting these adjustments in the overall RR. Two tiers are used: tier 1 for the low-quality studies, and tier 2 for the high-quality studies. The result changed from RR = 1.16 (95% CI 0.98 to 1.37), when there was no adjustment for augmentation, to RR =1.01 (95% CI 0.80 to 1.27) when publication was adjusted for using a monotonic weight function. Hence, adjustment had a considerable influence on this analysis.

It should be noted that the issue of varying quality of the included studies in a meta-analysis is a real one[349] and another report in this series deals with this issue further.[350] It should be stressed that this method does not adjust the individual or the pooled estimate for biases introduced by poor study design.

Final remarks on selection models

Many different approaches to weight functions have been outlined here. However, very few practical examples have been published that use such methods, and even fewer compare the results obtained by different selection models. In the discussion of Givens and co-workers' study,[295] although talking generally about selection models, DuMouchel observes that:

> "all attempts to assess publication bias beyond simple graphs like the funnel plot seem to involve a tour de force of modelling, and as such they are bound to run up against resistance from those who are [sceptical about] statistical modeling …"

Unfortunately there are, as yet, no software packages that facilitate this sort of analysis routinely; only when they are written will these methods be useable by the majority of researchers who carry out systematic reviews. It should be noted that an alternative method for adjusting for publication bias, the method of 'trim and fill' described below, is much easier to implement (although the availability of a software routine would also greatly facilitate its implementation). Preliminary analyses suggest that it may even outperform the most complex of the selection models. If further exploration confirms this finding, then it may make the selection model approach all but redundant.

Another point worth noting is that, in the discussion of Iyengar and Greenhouse's study,[290] Laird and colleagues commented on the similarity of these methods to those utilised in dealing with missing data in sample surveys: "Since the sample survey literature on handling non-response is extensive, we feel that many of the approaches developed for sample surveys can be used with advantage in the meta-analysis setting." This may be a methodology whose potential in the context of meta-analysis is not yet fully realised.

Other proposed methods for assessing/modelling publication bias

An array of methods other than selection models of the form described in the previous sections have been developed for assessing/modelling publication bias. These are discussed below.

The sensitivity approach of Copas

Copas[296] presents a method for adjusting for publication bias based on a procedure described by Copas and Li.[351] Using this method, the process of study selection is assumed to be described by a separate regression model with residuals that are correlated with study outcome. A likelihood approach is taken but the model cannot be fully identified without strong and unverifiable assumptions, so a sensitivity procedure based on an overall probability of study selection is adopted. A random effects meta-analysis model is used, together with a separate selection equation with a single correlation parameter (ρ) linking selection to outcome. The method is initially described using a test statistic as outcome; however, a modification allowing log RRs to be used is also given. The author also notes that an extension to allow the inclusion of covariates in the model would be relatively straightforward.

A brief exposition of the model is provided below, although this section can be skipped with no loss of continuity. The model is notated:

$$y = \mu + (\tau^2 + n^{-1})^{1/2}\varepsilon \qquad \text{equation 10}$$

and

$$z = \gamma_0 + \gamma_1 n^{1/2} + \delta \qquad \text{equation 11}$$

where τ^2 is the between-study variance and, in the example described, n^{-1} is the within-study variance, ε and δ are random residuals with a bivariate normal distribution with both means 0,

both variances 1, and correlation ρ. Equation 10 is a linear model describing the results of a population of studies that have been or could be carried out in this area, and μ is the main parameter of interest. However, only studies with positive values of z are available to the meta-analyst. This implies that the model for the actual studies in a meta-analysis is given by the conditional distribution of y given $z > 0$. If $\rho = 0$, this is the same as the ordinary distribution of y, but when $\rho \neq 0$ it models the selection bias; for example, if $\rho > 0$, selected studies will have $z > 0$ and hence are more likely to have large δ. Because δ is correlated with ε, this also tends to be large, leading to a positively biased value for y. Copas comments that the way n enters (equation 11) is a little arbitrary.

The probability of selection is given by:

$$P(z > 0 \backslash n) = \phi(\gamma_0 + \gamma_1 n^{1/2}) \qquad \text{equation 12}$$

where ϕ is the standard normal cumulative distribution function. This indicates that large studies are more likely to be selected than small studies. The average of the selected ys is:

equation 13

$$E(y \backslash z > 0, n) = \mu + \rho(\tau^2 + n^{-1})^{1/2} \lambda(\gamma_0 + \gamma_1 n^{1/2})$$

where λ is Mill's ratio of the standard normal density to the cumulative distribution function Φ. Unfortunately, as there are so many parameters that need estimating in this model, a sensitivity approach is more feasible than attempting to estimate the full model. Contour plots are presented in the original article, which assess changes in the pooled mean over values of the other parameters that define the selection model.

Applications of the method of Copas

An example from the rehabilitative treatments literature, specifically whether imprisonment reduces the likelihood that an offender will commit further offences, is presented.[296] Design quality is important when interpreting these studies because they range from randomised trials to informal research methods of "doubtful scientific value"; it is taken into account in the heterogeneity part of the model. When the model is applied to the data, the conclusions of the original meta-analysis stand, but the strength of the evidence is greatly reduced.

A further medical example, re-assessing the meta-analysis of Yusuf and co-workers,[148] which investigated the use of intravenous magnesium for patients with suspected acute myocardial infarction, is considered. This meta-analysis came to the conclusion that the use of magnesium was beneficial; however, 2 years later the results of a large trial were published, which contradicted this finding. It has been suggested that publication bias was the cause of the meta-analysis producing a result conflicting with the large trial. Re-modelling the data used by Yusuf and colleagues, taking into account heterogeneity and potential publication bias, produced a non-significant treatment effect over a wide array of values for parameters in the model.

The "trim and fill" method of Taylor and Tweedie

This is a new rank-based data augmentation technique, which formalises the use of funnel plots, and estimates and adjusts for the numbers and outcomes of missing studies.[353] There are two steps to the approach. First the number of missing studies is estimated using methods based on symmetry assumptions; and, secondly, missing values are imputed into the dataset, hence the name "trim and fill". An adjusted pooled effect size can be obtained from the augmented dataset. A more detailed description of this procedure is described below.

The authors note that previous methods of adjusting for publication bias (i.e. see selection models above) are complex and highly computer intensive to run; indeed, Dear and Dobson, in a comment on Givens and colleagues' article,[295] state that "previous methods have not been much used...[and]...the value of any new statistical methodology depends, in part, on the extent to which it is adopted".

In contrast, this method is both conceptually easy and simpler to implement than any of the previous methods described to adjust a meta-analysis for publication bias. It is non-parametric and relies only on symmetry assumptions, which are satisfied by both fixed and random effect meta-analysis models. The key assumption of the method is that it is the most extreme negative studies that have not been published.

Three different estimators for the number of missing studies are derived in the report.[352] The properties of these estimators are investigated via simulation studies. It is generally concluded that they work well in all but very extreme cases. An iterative "trim and fill" algorithm is described. Put non-technically, the estimated number of missing studies is "trimmed" from the asymmetric outlying

part of the funnel plot. Then the symmetrical remainder is used to estimate the "true centre" of the funnel. Then, the funnel plot is "filled" by replacing the "trimmed" studies and their estimated missing counterparts. This last stage is necessary for the variance of the pooled estimate to be calculated correctly. A subtlety of this approach that needs highlighting is that it assumes that studies are suppressed, not published under a scenario where it is the outcome magnitude, and not the *p*-value, which determines the chance of publication. This is the truncation that is picked up by "eye-balling" a funnel plot.

In a second report,[282] Duval and Tweedie further examine the properties of this approach.

1. Consideration of which of the three estimates of the number of missing studies have better mean squared error properties: They conclude that two methods are clearly better than the third, but there are values of the number of observed and missing studies for which each is better than the other (details of each estimator and ranges for which each appears optimal are given in the report).
2. Use of the distributional properties of the estimators to formulate tests for the existence of publication bias: These are then compared with the methods of Begg and Mazumdar,[288] and with Egger and co-workers.[44] The new test would appear to be quite powerful if there are more than five to six missing studies.

Simulation studies, where the suppression mechanism is known, suggest that trim and fill performs well, and that limited comparisons with selection models indicate comparable performance. In estimating the number of missing studies the authors consider their method to be more viable than the method of Gleser and Olkin.[283] It has the significant advantage of also being the most simple method. The authors do however stress that the main goal of their work should be seen as providing a method for sensitivity analyses rather than actually finding the values of missing studies, and that the method does seem to give good indications of which meta-analyses do not suffer from publication bias and which need to be evaluated much more carefully.

Five comparative applications of the method

This method has been illustrated by several examples that are particularly enlightening because they compare the trim and fill method with the method of Gleser and Olkin,[283] which also estimates the number of missing studies (see below), and the tests of Begg and Mazumdar,[288] and Egger and co-workers,[44] are also applied to the data.

1. In the first example, simulated data are used, so it is known that five studies were suppressed. The Gleser and Olkin method estimates that there are 11 missing studies, while the Begg test produces a *p*-value of 0.14 and the Egger test a *p*-value of 0.004 for publication bias. Hence, both tests would appear to detect a biased selection procedure because the test of Begg is known to have low power; thus a *p*-value of this magnitude should be taken as an indication of the possible existence of publication bias. The trim and fill method estimates the number of missing studies at between four and six (using the three different estimators proposed), and the pooled estimates and CIs produced are very close to those from the original complete data.
2. The second example considers the lung cancer and exposure to environmental tobacco smoke example considered previously (and described above) by Givens and colleagues.[295] Both Begg's and Egger's tests suggest that publication bias is present; however, Gleser and Olkin's method estimates the number of missing studies to be 0. The authors explain that this is because Gleser and Olkin's method can easily be influenced by just one study with a large *p*-value against the null hypothesis. In comparison, the trim and fill method estimates that between eight and nine studies are missing and produces an adjusted estimate and CI almost identical to that produced by the more complicated method of Givens and co-workers.[295]
3. The third example examines a dataset concerning teacher expectancy and later pupil performance on an IQ test. The methods of Gleser and Olkin yield three possible estimates of the number of missing studies, namely, 0, 59 or 82, which clearly vary hugely. The trim and fill method suggests that three studies are missing and, when these are imputed, the direction of the pooled effect is reversed (although it is non-significant). All test methods suggest that publication bias is present.
4. The fourth example examines the association between *Chlamydia trachomatis* and oral contraceptive use. This produces two quite different results for the number of missing studies, namely one and seven, depending on which of the three estimators is used, while Gleser and Olkin's method estimates 24. When the adjustment is made, the pooled result is reduced by 12%.

5. The final example[353] uses a dataset from the 1998 Cochrane Database of Systematic Reviews and investigates the effect of using gangliosides in reducing case fatality in acute stroke. A funnel plot of the 11 studies included in this meta-analysis presents "classic" publication bias symptoms with no studies plotted in the bottom left hand corner of the funnel. Indeed, this could be considered an extreme example. After the plot was trimmed and filled the pooled OR reduced from 1.11 (95% CI 0.88 to 1.13) to 1.01 (95% CI 0.82 to 1.26).

The impact of applying trim and fill to a collection of meta-analyses

An investigation to assess the possible magnitude of the impact that publication bias can have on the results and conclusions of a meta-analysis or systematic review has been undertaken using the trim and fill method.[354] Forty-eight meta-analyses included in the Cochrane Database of Systematic Reviews were examined. The trim and fill method estimated that 26 of these (54%) had at least one study missing when using a fixed effects model; in ten of these, the level of publication bias was significant at the 5% level. In four cases, statistical inferences regarding the intervention effect changed after adjustment of the overall estimate. Hence, this study suggests that a moderate level of publication bias exists in the medical literature.

Models that estimate the number of unpublished studies

Rosenthal's fail-safe N method for estimating the number of unpublished studies has been discussed in chapter 7.[5] Methods suggested by Gleser and Olkin,[283] by Eberly and Casella,[284] and by Ashworth and colleagues[355] are discussed below.

Methods suggested by Gleser and Olkin

The fail-safe N estimated by using Rosenthal's method is not necessarily related to the actual number of unpublished studies. Using the *p*-values reported in the published studies, Gleser and Olkin[283] proposed two general models for estimating the number of unpublished studies. They also provided methods for estimating the lower confidence bound for the estimated number of unpublished studies. The models are based on the following three basic assumptions.

1. All studies, published and unpublished, are mutually statistically independent.
2. The *p*-value in each study is based on a continuous test statistic.
3. The null hypothesis, H_0, is true.

A simple model for *p*-values is based on the assumption that the studies we observe are those that have the smallest *p*-value. Then the unbiased estimator of the number of unpublished studies (N_1) is:

$$N_1 = \frac{K(1 - p_{(k)}) - 1}{p_{(k)}} \qquad \text{equation 14}$$

where K is the number of published studies, and $p(k)$ is the largest *p*-value in observed studies.

The simple model can be more realistically generalised by assuming that K published studies consisted of m smallest *p*-values and n *p*-values that are a random sample from the $n + N_2$ largest *p*-values. Then an unbiased estimator of the number of unpublished studies is:

$$N_2 = \frac{m - 1}{p_{(m)}} - K \qquad \text{equation 15}$$

where $p_{(m)}$ is the largest in m smallest *p*-values.

The number of unpublished studies can also be estimated by a selection model.[283] The calculation needs first to assert the conditional probability (P_{ab}) that a *p*-value will be reported when it lies in the interval (a,b), and then $r = P_{ab} / (b - a)$. For example, it may be assumed that all *p*-values between 0.00 and 0.05 will be published. That is, $a = 0.00$, $b = 0.05$, and $P_{ab} = 1.00$; therefore $r = 1.00 \times (0.05 - 0.00) = 0.05$. Then the unbiased estimator of the number of unpublished studies is:

$$N_3 = \frac{d}{r} - k \qquad \text{equation 16}$$

where d is the number of observed *p*-values in the interval (a,b).

The "simple" method of equation 2 in this article gives results that are contradictory with those of the trim and fill method. Duval and Tweedie[282] suggest that this is largely due to the lack of robustness of Gleser and Olkin's method, both to isolated negative values (leading to the zero estimates) and in its heavy dependence on the null hypothesis.

Applying these methods to the example in *Figure 9*, the estimated number of unpublished

studies is $N_1 = 3$ by using the simple model; $N_2 = 4$ by assuming there were six smallest p-values ($m = 6$ and $p_{(m)} = 0.36$) and four p-values were a random sample from p-values that are greater than 0.36; and $N_3 = 12$ by using the selection model. (For the selection model, it was assumed that 95% of p-values are published when they lie between 0.00 and 0.05 (i.e. $r = 0.045$)).

Method of Eberly and Casella

Eberly and Casella[284] present a model for estimating the total number of studies carried out, both published and unpublished, dependent on the probability of publication. A simple selection model is used, where all studies significant at level α are published, while non-significant studies are published with probability p. This method varies from those using weight functions, however. Metropolis simulation and Gibbs sampling techniques[356] are used to generate random samples from the distribution of the total number of studies and explore how this changes as p varies. A β distribution is placed on the probability of publication to take into account the likely variation from field to field, from journal to journal, and from year to year. A simulation study examining the performance of the method is presented in the original report. Drawbacks of the method are that: it requires all studies considered to be roughly the same size; and only the count of the number of significant studies obtained is considered. This leads to a loss of information because individual p-values are not used.

Application of the method

The method is illustrated with a meta-analysis on studies of lead exposure and IQ levels in children, where the outcome IQ was measured on a continuous scale. The estimated number of unseen studies generated by this method were compared with Rosenthal's file drawer approach.[5] In this way one can assess if the estimated number of missing studies could change the conclusions of the meta-analysis. In this example, 5–11 studies were estimated as missing, whereas the file drawer estimate indicated that 16 would be needed to overturn the significant result, which suggests that the original conclusions are probably valid, if weakened. The author notes however that Rosenthal's method produces strictly *ad hoc* estimates and caution should always be used in interpreting the results.

Null-K method suggested by Ashworth and co-workers

Validity generalisation approaches involve combining correlation coefficients across studies and correcting these cumulative distributions for statistical artefacts, for estimating the true corrected correlations and the standard deviation of the true correlations. A 10% lower credibility value within the distribution of true correlations is often used to estimate the risk of having a true correlation that is too small to have practical utility.[355]

Ashworth and colleagues[355] suggested a 'null-K' method for assessing the vulnerability of validity generalisation results to unrepresented or missing studies. The null-K is the number of null unrepresented studies (with zero effect size and zero variation) required to bring the estimated 10% lower credibility value for a combined set of studies down to zero. The smaller the null-K, the greater the threat of potential publication bias to a positive validity generalisation finding.

The null-K method was tested by using 103 validity generalisation findings and the results were compared with that of Rosenthal's fail-safe N method.[355] It was found that the estimated null-Ks were much smaller than the estimated fail-safe Ns. Concerning the interpretation of estimated null-K, Ashworth and co-workers[355] suggested that "it is up to the individual researcher doing cumulative research to comment on the probability of the existence of unrepresented studies or the likelihood of a number of future studies capable of changing the scientific conclusion."

Estimating results of an unpublished study

Sugita and colleagues suggested a method for estimating the OR of an unpublished study in meta-analysis.[357,358] A summary OR was estimated by combining the estimated OR of the unpublished study with that of published studies. This method is based on the following assumptions.

1. A hypothetical study with the lnOR X_H and SE $1/\sqrt{W_H}$ has not been published.
2. The lnORs from all studies, published and unpublished, are distributed normally.

If there are K published studies, the lnOR and its SE can be estimated for each of the K studies: $lnORi = X_i$ and $\mathrm{SE}(lnORi) = 1/\sqrt{w_i}$, $(i = 1,...K)$. Next, a moment method using the sample second, third and fourth central moments (m_2, m_3 and m_4) of the normal distribution is used to obtain a set of simultaneous equations for estimating X_H and w_H:

$$m_3 = 0$$
$$m_4 - 3m_2^2 \approx 0$$

where

$$m_j = \frac{\sum w_i (X_i - X_A)^j + w_H (X_H - X_A)^j}{\sum w_i + w_H} \quad \text{equation 17}$$

and

$$X_A = \frac{\sum w_i X_i + w_H X_H}{\sum w_i + w_H} \quad \text{equation 18}$$

Based on the estimated X_H and w_H for the hypothetical unpublished study, the summary lnOR and its SE for all studies, published and unpublished, is: $lnOR_A = X_A$ and $\mathrm{SE}(lnOR_A) = 1/\sqrt{w_A} = 1/\sqrt{(\sum w_i + w_H)}$.

This method can be used to estimate the magnitude of the treatment effect in an unpublished study. It assumes that results across individual studies are homogeneous enough to use the fixed effects model for quantitative combination. Therefore, the method cannot be used when the random effects model has to be used. More importantly, it seems unrealistic to assume that only one study is unpublished.

Discussion and summary

Research into approaches for analytically assessing/adjusting a meta-analysis for publication bias has been an area of high activity over recent years, with many of the methods currently newly developed only described in unpublished reports. However, there is far from agreement among researchers that adjusting a meta-analysis for publication bias is a valid and sensible thing to do. Although several extensive simulation studies have been carried out, validation is difficult as one cannot usually identify the unpublished studies. Several of the authors consider that adjustment methods should be used as a form of sensitivity analysis only. This is an approach the authors of this report currently favour. If the results of a meta-analysis change dramatically after using one or a combination of the methods described here, caution is needed when drawing conclusions from the results obtained. In some instances, a funnel plot may display "classic" publication bias symptoms; however, even if the missing studies had been included, the result could still be very similar, as it is usually only the smallest studies that are not published. If this can be demonstrated through the use of analytical methods then this is valuable information; the meta-analysis of Linde and colleagues[298] may be a good example of this.

As for addressing the question of which of the methods is best to use, currently there is no clear answer; probably the use of a combination of methods is appropriate. However, the trim and fill method appears to provide sensible results and in the examples described above, seems to outperform several alternative methods in many circumstances. No in-depth comparison of its results with those of selection models has been carried out, and the authors of this report consider that this is necessary before one considers the other approaches to be obsolete. It remains to be seen how useful and appropriate these methods are in practice, and whether they move from experimental to more mainstream methodology.

Further research

It has been noted that publication bias affects the results of a random effects meta-analysis to a greater extent than those of a fixed effect one. Since on the whole the random effect approach has been recommended over that of fixed effects in a number of reports,[344,345] further exploration of how this affects the robustness of results is needed.

Begg, in the discussion of Givens and co-workers' study,[295] suggested that publication bias could artificially inflate between-study heterogeneity. The authors are not aware of any empirical evidence of this and it could be argued that removing a proportion of studies at the extremity of a distribution would actually artificially deflate the between-study variance. Either way, the impact of publication bias on the estimate of the between-study variance needs addressing. (It must be noted that methods such as trim and fill do address this issue by imputation to adjust the estimate of the between-study variance.)

Further comparisons of the performance of different methods is needed. Many of the examples used to illustrate the methods so far have not been from medicine or related areas, and performance of the methods by using such scales as the lnOR needs to be addressed.

If methods for adjusting meta-analysis are going to be taken up for routine use, then the development of software is required to implement these methods. Ideally, a single piece of software should be developed that can implement several of the methods

and allow an assessment of the sensitivity of the result to the method used to adjust. Bayesian selection models could perhaps be formulated in a program such as BUGS,[359] although previously the method of Givens and co-workers[295] could not be implemented in this package (Tweedie R, Division of Biostatistics, University of Minnesota: personal communication, 1998).

Health Technology Assessment panel membership

This report was identified as a priority by the Methodology Group.

Acute Sector Panel

Current members

Chair:
Professor Francis H Creed
University of Manchester

Professor Clifford Bailey
University of Leeds

Ms Tracy Bury
Chartered Society of Physiotherapy

Professor Collette Clifford
University of Birmingham

Dr Katherine Darton
M.I.N.D.

Mr John Dunning
Papworth Hospital, Cambridge

Mr Jonathan Earnshaw
Gloucester Royal Hospital

Mr Leonard Fenwick
Freeman Group of Hospitals, Newcastle-upon-Tyne

Professor David Field
Leicester Royal Infirmary

Ms Grace Gibbs
West Middlesex University Hospital NHS Trust

Dr Neville Goodman
Southmead Hospital Services Trust, Bristol

Professor Mark Haggard
MRC Institute of Hearing Research, University of Nottingham

Professor Robert Hawkins
University of Manchester

Dr Duncan Keeley
General Practitioner, Thame

Dr Rajan Madhok
East Riding Health Authority

Dr John Pounsford
Frenchay Hospital, Bristol

Dr Mark Sculpher
University of York

Dr Iqbal Sram
NHS Executive, North West Region

Mrs Joan Webster
Consumer member

Past members

Professor John Farndon*
University of Bristol

Professor Senga Bond
University of Newcastle-upon-Tyne

Professor Ian Cameron
Southeast Thames Regional Health Authority

Ms Lynne Clemence
Mid-Kent Health Care Trust

Professor Cam Donaldson
University of Aberdeen

Professor Richard Ellis
St James's University Hospital, Leeds

Mr Ian Hammond
Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris
Churchill Hospital, Oxford

Dr Gwyneth Lewis
Department of Health

Mrs Wilma MacPherson
St Thomas's & Guy's Hospitals, London

Dr Chris McCall
General Practitioner, Dorset

Professor Alan McGregor
St Thomas's Hospital, London

Professor Jon Nicholl
University of Sheffield

Professor John Norman
University of Southampton

Professor Michael Sheppard
Queen Elizabeth Hospital, Birmingham

Professor Gordon Stirrat
St Michael's Hospital, Bristol

Dr William Tarnow-Mordi
University of Dundee

Professor Kenneth Taylor
Hammersmith Hospital, London

* Previous Chair

continued

continued

Diagnostics and Imaging Panel

Current members

Chair:
Professor Mike Smith
University of Leeds

Dr Philip J Ayres
Leeds Teaching Hospitals NHS Trust

Dr Paul Collinson
St George's Hospital, London

Dr Barry Cookson
Public Health Laboratory Service, Colindale

Professor David C Cumberland
University of Sheffield

Professor Adrian Dixon
University of Cambridge

Mr Steve Ebdon-Jackson
Department of Health

Mrs Maggie Fitchett
Association of Cytogeneticists, Oxford

Dr Peter Howlett
Portsmouth Hospitals NHS Trust

Professor Alistair McGuire
City University, London

Dr Andrew Moore
Editor, *Bandolier*

Dr Peter Moore
Science Writer, Ashtead

Professor Chris Price
London Hospital Medical School

Dr William Rosenberg
University of Southampton

Mr Tony Tester
South Bedfordshire Community Health Council

Dr Gillian Vivian
Royal Cornwall Hospitals Trust

Dr Greg Warner
General Practitioner, Hampshire

Past members

Professor Michael Maisey*
Guy's & St Thomas's Hospitals, London

Professor Andrew Adam
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Pat Cooke
RDRD, Trent Regional Health Authority

Ms Julia Davison
St Bartholomew's Hospital, London

Professor MA Ferguson-Smith
University of Cambridge

Dr Mansel Haeney
University of Manchester

Professor Sean Hilton
St George's Hospital Medical School, London

Mr John Hutton
MEDTAP International Inc., London

Professor Donald Jeffries
St Bartholomew's Hospital, London

Dr Ian Reynolds
Nottingham Health Authority

Professor Colin Roberts
University of Wales College of Medicine

Miss Annette Sergeant
Chase Farm Hospital, Enfield

Professor John Stuart
University of Birmingham

Dr Ala Szczepura
University of Warwick

Mr Stephen Thornton
Cambridge & Huntingdon Health Commission

Dr Jo Walsworth-Bell
South Staffordshire Health Authority

Methodology Group

Current members

Chair:
Professor Martin Buxton
Health Economics Research Group, Brunel University

Professor Doug Altman
ICRF/NHS Centre for Statistics in Medicine, University of Oxford

Dr David Armstrong
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Nicholas Black
London School of Hygiene & Tropical Medicine

Professor Ann Bowling
University College London Medical School

Dr Mike Clarke
UK Cochrane Centre, Oxford

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Dr Vikki Entwistle
University of Aberdeen

Professor Ewan Ferlie
Imperial College, London

Professor Ray Fitzpatrick
University of Oxford

Mrs Jenny Griffin
Department of Health

Professor Jeremy Grimshaw
University of Aberdeen

Dr Stephen Harrison
University of Leeds

Mr John Henderson
Department of Health

Professor Richard Lilford
R&D, West Midlands

Professor Theresa Marteau
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Henry McQuay
University of Oxford

Dr Nick Payne
University of Sheffield

Professor Maggie Pearson
NHS Executive North West

Dr David Spiegelhalter
Institute of Public Health, Cambridge

Professor Joy Townsend
University of Hertfordshire

Ms Caroline Woodroffe
Standing Group on Consumers in NHS Research

Past members

Professor Anthony Culyer*
University of York

Professor Michael Baum
Royal Marsden Hospital

Dr Rory Collins
University of Oxford

Professor George Davey Smith
University of Bristol

Professor Stephen Frankel
University of Bristol

Mr Philip Hewitson
Leeds FHSA

Mr Nick Mays
King's Fund, London

Professor Ian Russell
University of York

Professor David Sackett
Centre for Evidence Based Medicine, Oxford

Dr Peter Sandercock
University of Edinburgh

Dr Maurice Slevin
St Bartholomew's Hospital, London

Professor Charles Warlow
Western General Hospital, Edinburgh

* Previous Chair

Pharmaceutical Panel

Current members

Chair:
Professor Tom Walley
University of Liverpool

Dr Felicity Gabbay
Transcrip Ltd

Dr Peter Golightly
Drug Information Services,
NHS Executive Trent

Dr Alastair Gray
Health Economics
Research Centre,
University of Oxford

Professor Rod Griffiths
NHS Executive
West Midlands

Mrs Jeanette Howe
Department of Health

Professor Trevor Jones
ABPI, London

Ms Sally Knight
Lister Hospital, Stevenage

Dr Andrew Mortimore
Southampton & SW Hants
Health Authority

Mr Nigel Offen
NHS Executive Eastern

Dr John Reynolds
The Oxford Radcliffe Hospital

Mrs Marianne Rigge
The College of Health,
London

Mr Simon Robbins
Camden & Islington
Health Authority, London

Dr Frances Rotblat
Medicines Control Agency

Dr Eamonn Sheridan
St James's University Hospital,
Leeds

Mrs Katrina Simister
National Prescribing Centre,
Liverpool

Dr Ross Taylor
University of Aberdeen

Past members

Professor Michael Rawlins*
University of Newcastle-
upon-Tyne

Dr Colin Bradley
University of Birmingham

Professor Alasdair
Breckenridge
RDRD, Northwest Regional
Health Authority

Ms Christine Clark
Hope Hospital, Salford

Mrs Julie Dent
Ealing, Hammersmith &
Hounslow Health Authority,
London

Mr Barrie Dowdeswell
Royal Victoria Infirmary,
Newcastle-upon-Tyne

Dr Tim Elliott
Department of Health

Dr Desmond Fitzgerald
Mere, Bucklow Hill, Cheshire

Professor Keith Gull
University of Manchester

Dr Keith Jones
Medicines Control Agency

Dr John Posnett
University of York

Dr Tim van Zwanenberg
Northern Regional
Health Authority

Dr Kent Woods
RDRD, Trent RO,
Sheffield

Population Screening Panel

Current members

Chair:
Professor Sir John
Grimley Evans
Radcliffe Infirmary, Oxford

Mrs Stella Burnside
Altnagelvin Hospitals Trust,
Londonderry

Mr John Cairns
University of Aberdeen

Professor Howard Cuckle
University of Leeds

Dr Carol Dezateux
Institute of Child Health,
London

Mrs Anne Dixon-Brown
NHS Executive Eastern

Professor Dian Donnai
St Mary's Hospital,
Manchester

Dr Tom Fahey
University of Bristol

Mrs Gillian Fletcher
National Childbirth Trust

Dr JA Muir Gray
National Screening
Committee, NHS Executive
Oxford

Professor Alexander Markham
St James's University Hospital,
Leeds

Dr Ann McPherson
General Practitioner,
Oxford

Dr Susan Moss
Institute of Cancer Research

Mr John Nettleton
Consumer member

Mrs Julietta Patnick
NHS Cervical
Screening Programme,
Sheffield

Dr Sarah Stewart-Brown
Health Service Research Unit,
University of Oxford

Past members

Dr Sheila Adam*
Department of Health

Professor George Freeman
Charing Cross & Westminster
Medical School, London

Dr Mike Gill
Brent & Harrow
Health Authority

Dr Anne Ludbrook
University of Aberdeen

Professor Theresa Marteau
Guy's, King's &
St Thomas's School of
Medicine & Dentistry,
London

Professor Catherine Peckham
Institute of Child Health,
London

Dr Connie Smith
Parkside NHS Trust,
London

Ms Polly Toynbee
Journalist

Professor Nick Wald
University of London

Professor Ciaran Woodman
Centre for
Cancer Epidemiology,
Manchester

* Previous Chair

continued

continued

Primary and Community Care Panel

Current members

Chair:
Dr John Tripp
Royal Devon & Exeter Healthcare NHS Trust

Mr Kevin Barton
East London & City Health Authority

Professor John Bond
University of Newcastle-upon-Tyne

Dr John Brazier
University of Sheffield

Ms Judith Brodie
Cancer BACUP

Mr Shaun Brogan
Ridgeway Primary Care Group, Aylesbury

Mr Joe Corkill
National Association for Patient Participation

Dr Nicky Cullum
University of York

Professor Pam Enderby
University of Sheffield

Dr Andrew Farmer
Institute of Health Sciences, Oxford

Dr Jim Ford
Department of Health

Professor Richard Hobbs
University of Birmingham

Professor Allen Hutchinson
University of Sheffield

Dr Aidan MacFarlane
Independent Consultant

Professor David Mant
Institute of Health Sciences, Oxford

Dr Chris McCall
General Practitioner, Dorset

Dr Robert Peveler
University of Southampton

Professor Jennie Popay
University of Salford

Dr Ken Stein
North & East Devon Health Authority

Past members

Professor Angela Coulter*
King's Fund, London

Professor Martin Roland*
University of Manchester

Dr Simon Allison
University of Nottingham

Professor Shah Ebrahim
Royal Free Hospital, London

Ms Cathy Gritzner
King's Fund, London

Professor Andrew Haines
RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks
Oxfordshire Health Authority

Mr Edward Jones
Rochdale FHSA

Professor Roger Jones
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Mr Lionel Joyce
Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp
London School of Economics & Political Science

Dr Phillip Leech
Department of Health

Professor Karen Luker
University of Liverpool

Dr Fiona Moss
Thames Postgraduate Medical & Dental Education

Professor Dianne Newham
King's College London

Professor Gillian Parker
University of Leicester

Dr Mary Renfrew
University of Oxford

Ms Hilary Scott
Tower Hamlets Healthcare NHS Trust, London

* Previous Chair

National Coordinating Centre for Health Technology Assessment, Advisory Group

Current members

Chair:
Professor John Gabbay
Wessex Institute for Health Research & Development

Dr Sheila Adam
Department of Health

Professor Nicholas Black
London School of Hygiene and Tropical Medicine

Professor Martin Buxton
Health Economics Research Group, Brunel University

Mr Harry Cayton
Alzheimer's Disease Society

Professor Angela Coulter
The King's Fund, London

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Professor Shah Ebrahim
MRC Health Services Research Collaboration, University of Bristol

Ms Lynn Kerridge
Wessex Institute for Health Research & Development

Professor Jos Kleijnen
NHS Centre for Reviews and Dissemination, University of York

Dr Ruairidh Milne
Wessex Institute for Health Research & Development

Ms Kay Pattison
Research & Development Directorate, NHS Executive

Professor James Raftery
Health Economics Unit, University of Birmingham

Professor Ian Russell
Department of Health Sciences & Clinical Evaluation, University of York

Dr Ken Stein
North & East Devon Health Authority

Professor Andrew Stevens
Department of Public Health & Epidemiology, University of Birmingham

Professor Kent Woods
Department of Medicine & Therapeutics, University of Leicester

Past member

Dr Paul Roderick
Wessex Institute for Health Research & Development