Health Technology Assessment 2005; Vol. 9: No. 26

Indirect comparisons of competing interventions

AM Glenny, DG Altman, F Song,
C Sakarovitch, JJ Deeks, R D'Amico,
M Bradburn and AJ Eastwood

July 2005

Health Technology Assessment
NHS R&D HTA Programme

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.

Indirect comparisons of competing interventions

AM Glenny,[1*] DG Altman,[2] F Song,[3]
C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2]
M Bradburn[2] and AJ Eastwood[4]

In collaboration with the International Stroke Trial Collaborative Group

[1] Cochrane Oral Health Group, Dental School, University of Manchester, UK
[2] Cancer Research UK, Centre for Statistics in Medicine, Wolfson College Annexe, Oxford, UK
[3] School of Medicine Health Policy and Practice, University of East Anglia, Norwich, UK
[4] Centre for Reviews and Dissemination, University of York, UK

*Corresponding author

Declared competing interests of authors: none

Published July 2005

This report should be referenced as follows:

Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, *et al*. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;**9**(26).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE and *Science Citation Index Expanded (SciSearch®)* and *Current Contents®*/Clinical Medicine.

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role 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 provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, consumer groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including consumers) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or designing a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a limited time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

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.

The research reported in this monograph was commissioned by the HTA Programme as project number 96/51/99. The contractual start date was in February 1999. The draft report began editorial review in May 2002 and was accepted for publication in November 2004. As the funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

M

Abstract

Indirect comparisons of competing interventions

AM Glenny,[1*] DG Altman,[2] F Song,[3] C Sakarovitch,[2] JJ Deeks,[2] R D'Amico,[2] M Bradburn[2] and AJ Eastwood[4]

In collaboration with the International Stroke Trial Collaborative Group

[1] Cochrane Oral Health Group, Dental School, University of Manchester, UK
[2] Cancer Research UK, Centre for Statistics in Medicine, Wolfson College Annexe, Oxford, UK
[3] School of Medicine Health Policy and Practice, University of East Anglia, Norwich, UK
[4] Centre for Reviews and Dissemination, University of York, UK
*Corresponding author

Objectives: To survey the frequency of use of indirect comparisons in systematic reviews and evaluate the methods used in their analysis and interpretation. Also to identify alternative statistical approaches for the analysis of indirect comparisons, to assess the properties of different statistical methods used for performing indirect comparisons and to compare direct and indirect estimates of the same effects within reviews.

Data sources: Electronic databases.

Review methods: The Database of Abstracts of Reviews of Effects (DARE) was searched for systematic reviews involving meta-analysis of randomised controlled trials (RCTs) that reported both direct and indirect comparisons, or indirect comparisons alone. A systematic review of MEDLINE and other databases was carried out to identify published methods for analysing indirect comparisons. Study designs were created using data from the International Stroke Trial. Random samples of patients receiving aspirin, heparin or placebo in 16 centres were used to create meta-analyses, with half of the trials comparing aspirin and placebo and half heparin and placebo. Methods for indirect comparisons were used to estimate the contrast between aspirin and heparin. The whole process was repeated 1000 times and the results were compared with direct comparisons and also theoretical results. Further detailed case studies comparing the results from both direct and indirect comparisons of the same effects were undertaken.

Results: Of the reviews identified through DARE, 31/327 (9.5%) included indirect comparisons. A further five reviews including indirect comparisons were identified through electronic searching. Few reviews carried out a formal analysis and some based analysis on the naive addition of data from the treatment arms of interest. Few methodological papers were identified. Some valid approaches for aggregate data that could be applied using standard software were found: the adjusted indirect comparison, meta-regression and, for binary data only, multiple logistic regression (fixed effect models only). Simulation studies showed that the naive method is liable to bias and also produces over-precise answers. Several methods provide correct answers if strong but unverifiable assumptions are fulfilled. Four times as many similarly sized trials are needed for the indirect approach to have the same power as directly randomised comparisons. Detailed case studies comparing direct and indirect comparisons of the same effect show considerable statistical discrepancies, but the direction of such discrepancy is unpredictable.

Conclusions: Direct evidence from good-quality RCTs should be used wherever possible. Without this evidence, it may be necessary to look for indirect comparisons from RCTs. However, the results may be susceptible to bias. When making indirect comparisons within a systematic review, an adjusted indirect comparison method should ideally be used employing the random effects model. If both direct and indirect comparisons are possible within a review, it is recommended that these be done separately before considering whether to pool data. There is a need to evaluate methods for the analysis of indirect comparisons for continuous data and for empirical research into how different methods of indirect

comparison perform in cases where there is a large treatment effect. Further study is needed into when it is appropriate to look at indirect comparisons and when to combine both direct and indirect comparisons. Research into how evidence from indirect comparisons compares to that from non-randomised studies may also be warranted. Investigations using individual patient data from a meta-analysis of several RCTs using different protocols and an evaluation of the impact of choosing different binary effect measures for the inverse variance method would also be useful.

Contents

List of abbreviations

ACE	angiotensin-converting enzyme
AIC	adjusted indirect comparison
AIIA	angiotensin II antagonist
ALT	alanine transaminase
AP	aerolised pentamidine
APSAC	anisoylated plasminogen streptokinase activator complex
ASA	aminosalicylate
ATC	Antiplatelet Trialists' Collaboration
CAD	coronary artery disease
CDSR	Cochrane Database of Systematic Reviews
Cefot-M	cefotaxime plus metronidazole
Cefur-M	cefuroxime plus metronidazole
CI	confidence interval
Co-A	co-amoxiclav
CRD	Centre for Reviews and Dissemination
CT	chemotherapy
ctrl	control
D	dapsone
DARE	Database of Abstracts of Reviews of Effects
DC	direct comparison
Dip	dipyridamole
DP	D-penicillamine
DVT	deep vein thrombosis
EE	ethinyl estradiol
ENRIS	Evaluating Non-Randomised Intervention Studies
ESRD	end-stage renal disease
FTT	Fibrinolytic Therapy Trialists
5-FU	5-fluorouracil
GI	gastrointestinal
GLMM	generalised linear mixed model
GORD	gastro-oesophageal reflux disease
H_2RA	H_2-receptor antagonist
IBS	irritable bowel syndrome
IC	indirect comparison
IPD	individual patient data
IST	International Stroke Trial
IVE	important vascular events
LMWH	low molecular weight heparin
MAP	mean arterial pressure
NA	not applicable
NNT	number needed to treat
NRT	nicotine replacement therapy
ns	not significant
NSAID	non-steroidal anti-inflammatory drug
OA	osteoarthritis
OGD	oesophagogastroduodenoscopy
OR	odds ratio
P	pyrimethamine
PE	pulmonary embolism
PEP	polyestradiol phosphate
PPI	proton pump inhibitor
RA	rheumatoid arthritis
RCT	randomised controlled trial
RemRR	remedication rate ratio
ResRR	response rate ratio
RPA	reteplase
RR	relative risk
RT	radiotherapy

continued

List of abbreviations *continued*

SD	standard deviation
SE	standard error
SEM	standard error of the mean
SK	streptokinase
SMX	sulfamethoxazole
SPID	sum of pain intensity difference
SSRI	selective serotonin reuptake inhibitor
SSZ	sulfasalazine
SWI	surgical wound infection
TMP	trimethoprim
TOTPAR	total pain relief
t-PA	tissue plasminogen activator
UFH	unfractionated heparin

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

The randomised controlled trial (RCT) is the most valid design for evaluating the relative efficacy of healthcare technology. However, many competing interventions have not been directly compared in RCTs and indirect methods have been commonly used in meta-analyses. Such indirect comparisons are subject to greater bias (especially selection bias) than head-to-head randomised comparisons, as the benefit of randomisation does not hold across trials. Therefore, it is essential to evaluate such bias that may lead to inaccuracies in the estimates of treatment effects and result in inappropriate policy decisions.

Objectives

The objectives of this study were:

- to survey the frequency of use of indirect comparisons in systematic reviews and evaluate the methods used in their analysis and interpretation
- to identify alternative statistical approaches for the analysis of indirect comparisons
- to assess the properties of different statistical methods used for performing indirect comparisons
- to carry out empirical work comparing direct and indirect estimates of the same effects within reviews.

Methods

The Database of Abstracts of Reviews of Effects (DARE) (1994 to March 1999) was searched for systematic reviews involving meta-analysis of RCTs that reported both direct and indirect comparisons, or indirect comparisons alone. A systematic review of MEDLINE (1966 to February 2001) and other databases was carried out to identify published methods for analysing indirect comparisons.

Study designs were created using data from the International Stroke Trial. Random samples of patients receiving aspirin, heparin or placebo in 16 centres were used to create meta-analyses, with half of the trials comparing aspirin and placebo and half heparin and placebo. Methods for indirect comparisons were used to estimate the contrast between aspirin and heparin. The whole process was repeated 1000 times and the results were compared with direct comparisons and also theoretical results.

Further detailed case studies comparing the results from both direct and indirect comparisons of the same effects were undertaken.

Results

Of the reviews identified through DARE that included meta-analyses of two or more RCTs, 31/327 (9.5%) included indirect comparisons. A further five reviews including indirect comparisons were identified through electronic searching. Few reviews carried out a formal analysis. Some reviews based analysis on the naive addition of data from the treatment arms of interest. Interpretation of indirect comparisons was not always appropriate.

Few methodological papers were identified. Some valid approaches for aggregate data that could be applied using standard software were found: the adjusted indirect comparison, meta-regression and, for binary data only, multiple logistic regression (fixed effect models only).

Simulation studies showed that the naive method is liable to bias and also produces over-precise answers. Several methods provide correct answers if strong but unverifiable assumptions are fulfilled. Four times as many similarly sized trials are needed for the indirect approach to have the same power as directly randomised comparisons.

Detailed case studies comparing direct and indirect comparisons of the same effect show considerable statistical discrepancies, but the direction of such discrepancy is unpredictable.

Conclusions

When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence

from good-quality RCTs should be used wherever possible. If little or no such evidence exists, it may be necessary to look for indirect comparisons from RCTs. The reviewer needs, however, to be aware that the results may be susceptible to bias.

When making indirect comparisons within a systematic review, an adjusted indirect comparison method should ideally be used using the random effects model. If both direct and indirect comparisons are possible within a review, it is recommended that these be done separately before considering whether to pool data.

Recommendations for research

There is a need for evaluation of methods for analysis of indirect comparisons for continuous data.

There is a need for empirical research into how different methods of indirect comparison perform in cases where there is a large treatment effect.

Further research is required to consider how to determine when it is appropriate to look at indirect comparisons and how to judge when to combine both direct and indirect comparisons. Research into how evidence from indirect comparisons compares to that from non-randomised studies may also be warranted.

Empirical investigations were based on one large, multicentre trial with a common protocol across each centre. It would be useful to repeat the investigations using individual patient data from a meta-analysis of several RCTs using different protocols.

The odds ratio was used as the measure of effect within this simulation study. Although logistic regression calls for the effect measure to be the odds ratio, it would be interesting to evaluate the impact of choosing different binary effect measures for the inverse variance method.

Chapter 1
Background

Well-designed randomised controlled trials (RCTs) generally provide the most reliable evidence of effectiveness as observed differences between the trial arms can, in general, be confidently attributed to differences in the treatment(s) being evaluated.[1,2] However, in many areas, available trials may not have directly compared the specific treatments or regimens of interest. A common example is where there is a class of several drugs, each of which has been studied in placebo-controlled RCTs, but there are no trials (or very few) in which the drugs have been directly compared with each other. For example, in a systematic review of antibiotic prophylaxis for preventing surgical wound infections after colorectal surgery, only a limited number of antibiotics or combinations of antibiotics (more than 70 regimens in total) were directly compared (head-to-head comparisons) in 147 randomised trials.[3] With the increasing use and development of meta-analytical techniques, comparisons of arms of different RCTs are being undertaken. Such indirect comparisons are subject to greater bias (especially selection bias) than head-to-head randomised comparisons, as the benefit of randomisation does not hold across trials. It is vital, therefore, to evaluate such biases that may lead to inaccuracies in the estimates of treatment effects and result in inappropriate policy decisions. By identifying the presence, magnitude and, if possible, methods to overcome such bias, the accuracy and interpretation of estimates of the effectiveness of health technologies will be enhanced. In addition, such evaluations may give an insight into the usefulness of indirect comparisons in cases where direct comparisons are impossible, for example, when examining the placebo effects of subcutaneous and oral medication.[4]

In a review exploring the types of evidence used to support the prescribing of one drug over another, McAlister and colleagues[5] suggest a hierarchy of evidence for grading studies that compare a drug with another of the same class (*Table 1*).

As expected, direct comparisons from head-to-head RCTs measuring clinically important outcomes (referring to long-term efficacy data) are at the top of the hierarchy of evidence at level 1, followed by head-to-head RCTs using validated surrogate outcomes (level 2). Comparisons made across placebo-controlled RCTs of different drugs are, however, also classified at level 2, despite the increased threats to validity due to likely differences between trials in terms of end-point definitions, inclusion criteria, patient characteristics, setting or baseline risk of outcomes. McAlister and colleagues[5] acknowledge that the strength of inference from such indirect comparisons is limited.

It could be argued that since the randomisation element of the RCT is not (or not fully) used during indirect comparison, such methods may have important implications on the use of data

TABLE 1 *Suggested levels of evidence for comparing the efficacy of drugs within the same class*

Level	Comparison	Study patients	Outcomes
1	Within a head-to-head RCT	Identical (by definition)	Clinically important
2	Within a head-to-head RCT	Identical (by definition)	Validated surrogate
2	Across RCTs of different drugs vs placebo	Similar or different (in disease status and risk factor status)	Clinically important or validated surrogate
3	Across subgroup analyses from RCTs of different drugs vs placebo	Similar or different	Clinically important or surrogate
3	Across RCTs of different drugs vs placebo	Similar or different	Unvalidated surrogate
4	Between non-randomised studies	Similar or different	Clinically important

Adapted from McAlister *et al*.[5]

from other types of non-randomised ('observational') study design.

Indirect comparisons are not only used to make comparisons between drugs in the same class or during subgroup analyses when comparing, for example, different dosages of the same drug. Comparisons between very different interventions, such as pharmacological interventions versus surgery, or between different classes of drugs have also been made using indirect comparisons. Examples of each type of comparison can be seen in the health research literature.

For example, an indirect comparison of different classes of drugs was undertaken in a meta-analysis of second-line drugs used to treat rheumatoid arthritis.[6] The drugs compared were antimalarial drugs, auranofin, injectable gold, methotrexate, D-penicillamine and sulfasalazine. Sixty-six trials examining the efficacy of second line drugs were included. For each drug the results were combined across treatment groups. Means for each drug treatment were generated by weighting each treatment group by size at the end of the trial. To compare drugs, analysis of variance was undertaken, weighted by treated group size, multiplied by study quality and adjusted for covariates shown to have significant association with each outcome. A fixed effects model was used. The outcomes of interest were the tender joint count, the erythrocyte sedimentation rate and grip strength. For each outcome, results showed that auranofin tended to be weaker than other second line drugs. No attempt, however, was made to provide data from direct comparisons.

An example of indirect comparisons being used to compare drugs within a specific class can be seen in the systematic review by Garg and Yusuf.[7] The authors of the review used indirect comparisons to evaluate the effects of angiotensin-converting enzyme (ACE) inhibitors on mortality and morbidity in patients with symptomatic congestive heart failure. Thirty-two placebo-controlled trials of ACE inhibitors, including 7105 patients, were used to make adjusted indirect comparisons of estimates of effect for the different agents evaluated. The authors state that "Similar benefits were observed with several different ACE inhibitors, although the data were largely based on enalapril maleate, captopril, ramipril, quinapril hydrochloride, and lisinopril".

Indirect comparisons of the effect of different doses of a drug can be illustrated with the review by the Homocysteine Lowering Trialists' Collaboration.[8] The review included only randomised trials (with an untreated control group) that assessed the effects of different doses of folic acid, with or without the addition of vitamin B_{12} or B_6, on blood homocysteine concentrations. Twelve trials, with data on 1114 patients, were included in the review. Trials were grouped according to the folic acid regimen used (<1, 1–3 or >3 mg daily). The proportional reductions in blood homocysteine in the treated groups compared with the control groups were evaluated by an analysis of covariance which estimated the difference in the post-treatment, log-transformed homocysteine values after adjustment for baseline homocysteine levels. This model was extended to allow the extent of this adjustment to vary between studies according to factors such as folic acid dose, additional vitamin B_6 or B_{12}, age, gender or duration of treatment. After adjusting for pretreatment blood concentrations of homocysteine and folate, there was no significant difference between daily doses. Such indirect comparison of results from subgroup analyses is not uncommon in systematic reviews and meta-analyses. Boersma and colleagues[9] evaluated the effect of delayed thrombolytic treatment following acute myocardial infarction and short-term mortality. They included 22 randomised trials of over 50,000 patients. The trials were grouped according to time to fibrinolytic therapy (0–1, ≥ 1–2, ≥ 2–3, ≥ 3–6, ≥ 6–12, ≥ 12–24 hours) rather than dosage. The results of the subgroup analysis were used to conclude that the "beneficial effect of fibrinolytic therapy is substantially higher in patients presenting within two hours after symptom onset compared to those presenting later".

The above examples illustrate both unadjusted (naive) and adjusted indirect comparisons. In the naive approach[6] data are pooled across treatment arms, ignoring the fact that the studies are RCTs (and discarding data from some treatment groups). Given that comparisons are being made across trial arms in the naive indirect comparisons, negating the randomised nature of the trial, the exclusion of non-randomised or observational studies can be questioned. Such methods may have important implications for the use of data from other types of non-randomised (observational) study design.

In the adjusted indirect comparison, the comparison of the interventions of interest is adjusted by the results of their direct comparison with a common control group (e.g. placebo), partially using the strength of the RCT. The

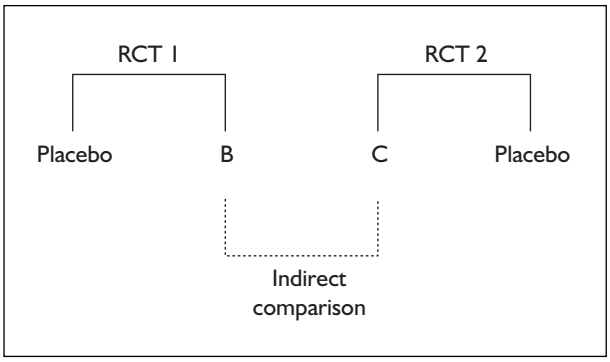

FIGURE 1 *Indirect comparison*

methods used to undertake such adjusted indirect comparisons vary. In the simplest case, one may be interested in comparing two interventions B and C. Indirect evidence can be obtained from RCTs of either B or C versus a common comparator, perhaps placebo (RCT 1 and RCT 2, *Figure 1*). One could use the results from one or more trials to estimate the effect of each intervention relative to control. An indirect measure of the relative efficacy of B and C could be made subjectively or using some statistical procedure. This information could be combined with data from any direct B versus C comparison, again either qualitatively or statistically. In more complex situations, one may be interested in trying to estimate simultaneously the (relative) effectiveness of several treatments (e.g. various β-blockers and other preventive treatments for patients who have had a myocardial infarction, or numerous analgesic drugs used in a variety of pain relief settings).

Several ad hoc approaches have been adopted to handle such situations. For example, in the case where there are a large number of placebo-controlled trials of various drugs a common approach is to carry out separate meta-analyses of the trials of each drug. The estimated effects are then compared explicitly or implicitly, ignoring the fact that the studies (or the patients in them) may not be strictly comparable. This procedure is like producing a league table, and results of this type appear increasingly in Bandolier[10] and elsewhere.[11] Such comparisons may be of little use if tables are ranked according to number needed to treat (NNT) estimates in which no-treatment groups, placebo-controlled trials and head-to-head comparisons are included.[12]

Meta-analyses using both randomised and non-randomised studies have also been undertaken. For example, in a review of thromboprophylaxis after total hip replacement, Murray and colleagues analysed treatment arms from both controlled and uncontrolled studies.[13] Such indirect comparisons between non-randomised groups are made on the assumption that these groups are similar across the different studies. However, this may not be a reasonable assumption, so analyses of this kind are open to several sources of bias. For example, as the efficacy of a treatment may vary among subpopulations of patients, differences in baseline characteristics between groups within different trials (variations in case-mix) may lead to biased estimates of treatment effect. Even when patient characteristics are similar, other aspects may vary between trials, such as ancillary treatment or other aspects of patient care (the actual treatment may vary too).

There is a need to investigate the properties of such procedures and to address the ways in which such meta-analyses can best be carried out. Most obviously, if trials have all used a common control intervention (maybe placebo) then there is the potential to use the control groups as a standardising factor. Bucher and colleagues[14] presented a model for undertaking indirect comparisons that preserves the randomisation of the originally assigned patient groups. Their model was tested using a meta-analysis of RCTs comparing two experimental regimens against the standard regimen for the prevention of *Pneumocystis carinii* pneumonia in patients with HIV infection. Trials providing a direct estimate of the relative effectiveness of the two experimental drugs were also identified, and an overall estimate of effect of these trials was calculated. Both the indirect and direct comparisons favoured trimethoprim–sulfamethoxazole over dapsone/pyrimethamine, but the magnitude of difference was less in the direct comparison. The odds ratio (OR) from the indirect comparison was 0.37 [95% confidence interval (CI) 0.21 to 0.65] favouring trimethoprim-sulfamethoxazole, compared with 0.64 (95% CI 0.45 to 0.90) for the direct comparison using the fixed effects model. Using the random effects model the odds ratio from the direct comparison was 0.43 (95% CI 0.21 to 0.89), which was similar to that from the indirect comparison. The model presented may protect against some of the biases that arise through the use of indirect comparisons. The authors concluded that only where direct comparisons are unavailable should indirect comparison meta-analysis be carried out. In such cases the limitations of such procedures should be examined thoroughly. However, the approach clearly needs further evaluation.

Indirect comparisons are being used in systematic reviews to evaluate the relative effectiveness of alternative interventions even though such indirect comparisons may be less accurate than head-to-head randomised comparisons. It is vital, therefore, to evaluate the properties of different statistical approaches to indirect comparisons to ensure that inaccuracies in the estimates of treatment effects do not result in inappropriate policy decisions.

Chapter 2

Research questions

The aims of the project were:

- to survey the frequency of use of indirect comparisons in systematic reviews and evaluate the methods used in their analysis and interpretation
- to identify alternative statistical approaches for the analysis of indirect comparisons
- to assess the properties of different statistical methods used for performing indirect comparisons
- to carry out empirical work comparing direct and indirect estimates of the same effects within reviews.

To achieve these aims the project involved a review of the literature and methodological and empirical investigations.

Chapter 3 presents a survey of published indirect comparisons. Chapter 4 contains a systematic review of the literature identifying different statistical approaches to the analysis of indirect comparisons. Chapters 5 and 6 present the methodological and empirical investigations. Detailed case studies are given in Chapter 7. Each chapter contains a description of the methods used, the results and a brief overview of the findings. Chapter 8 gives an overall discussion, drawing on the findings from Chapter 3–7, and the implications are discussed in Chapter 9.

Chapter 3

Indirect comparisons in published systematic reviews

This chapter summarises the findings of a survey of frequency of use of indirect comparisons in published systematic reviews.

Methods

Search strategy

To survey the frequency of use of indirect comparisons in published systematic reviews it was unnecessary to conduct a comprehensive search of the research literature. Rather, a careful search of databases of systematic reviews was performed. In addition, it was considered impossible to develop a search strategy to identify relevant published reviews.

Two databases were readily available which provide a source of meta-analyses: the Database of Abstracts of Reviews of Effects (DARE) [available through a number of sources, including the Centre for Reviews and Dissemination (CRD) website http://www.york.ac.uk/inst/crd/darehp.htm and the Cochrane Library] and the Cochrane Database of Systematic Reviews (CDSR) (on the Cochrane Library, available to all members of the NHS through the National Electronic Library for Health, www.nelh.nhs.uk). It was felt appropriate to focus initially on a search of DARE. DARE is a database of quality-assessed systematic reviews. The reviews are identified through regular searching of a number of electronic databases (including MEDLINE, CINAHL, Current Contents Clinical Medicine and BIOSIS), by handsearching key major journals and scanning grey literature. To be included on DARE, the reviews undergo a rigorous quality assessment and must meet set criteria with regard to the review question, the search strategy, validity assessment of the primary studies and the presentation and pooling of the primary studies. A hard copy of all reviews published on DARE (1994 to December 1998) was obtained, and each review screened according to the inclusion criteria (see below).

Inclusion criteria

All systematic reviews including at least one meta-analysis were assessed to see whether they used:

1. RCTs
2. indirect comparisons
3. direct comparisons.

The following types of meta-analysis were recorded:

- those incorporating elements 1 and 2, with the possibility of comparing them with other studies making direct comparisons of the same interventions. A note was made of whether the authors interpreted the results as if direct comparisons had been made
- those incorporating all three elements, to examine the differences in estimates of effect obtained using direct and indirect comparisons
- those incorporating elements 1 and 3, and providing sufficient data to try to undertake an indirect comparison of specific interventions.

All other systematic reviews were excluded.

Assessment of relevance

All systematic reviews listed on DARE were assessed using a piloted prescreening form (see Appendix 1) designed to allow for identification of articles meeting the above inclusion criteria and also those for the Evaluating Non-Randomised Intervention Studies (ENRIS) project.[15] Two reviewers independently assessed each article and disagreements were resolved by discussion. All data were managed using Microsoft Excel 97.

Data extraction

For all reviews assessed as relevant, data were extracted using a predefined data extraction form (Appendix 2). This process was undertaken independently by two reviewers and discrepancies were resolved through discussion. The results of the data extraction process were tabulated. The method used to carry out the indirect comparison and the appropriateness of the interpretation of results were noted. The interpretation was deemed appropriate if the findings of direct comparisons were given greater weight in the conclusions, or the potential biases associated with the findings of the indirect comparisons were discussed.

TABLE 2 *Reviews using both direct and indirect comparisons*

	No. of studies	Appropriate interpretation of results			Agreement between IC and DC (estimate of effect in same direction)		
		Yes	No	Uncertain	Yes	No	Uncertain
Adjusted IC	10	10	–	–	7	2	1
Naive IC	3	2	–	1	1	1	1

DC, direct comparison; IC, indirect comparison.

Results

In total, 734 systematic reviews were available on DARE for screening in March 1999. Of these, 327 included a meta-analysis of RCTs. The initial screening identified ten reviews that were coded as being definite examples of indirect comparisons and 46 reviews that were coded as possibly including an indirect comparison. A further 12 potentially relevant reviews were located through the electronic searches for the systematic review (Chapter 4).

After detailed assessment, 13 of these 68 reviews were identified as including both direct and indirect comparisons of competing interventions and 23 included indirect comparisons only. Thirty-one of the reviews were identified from the search of DARE and five from the electronic searches. Detailed summaries of the 36 included reviews appear in Appendix 3. The remainder of the reviews did not include suitable data and so were excluded (see Appendix 4).

Reviews including both direct and indirect comparisons

Thirteen of the identified systematic reviews used both direct and indirect comparisons.[16–28] Ten presented adjusted indirect comparisons, first estimating the pooled effects of each treatment arm against placebo or a control group and then comparing the two estimates. The interpretation of the results was appropriate in each case, with conclusions being based largely on the direct comparisons, or problems associated with indirect comparisons being discussed.[16–20,22,24,26–28] The other three reviews presented naive, unadjusted, indirect comparisons.[21,23,25] The interpretation of the results was classed as appropriate in two of the reviews, with emphasis given to results from direct comparisons when available.[21,25] There was some uncertainty surrounding the appropriateness of the interpretation of the results presented in the third review[23] (*Table 2*).

Each of these reviews is discussed in the following sections.

Adjusted indirect comparisons (see Appendix 3, Table 19)

Adjusted indirect comparisons were considered to have been used when estimates of effect were compared for interventions that had been evaluated using a common comparator (e.g. placebo). Three of the reviews undertaking an adjusted indirect comparison were conducted by the Antiplatelet Trialists' Collaboration and examined the effects of antiplatelet therapy on various outcomes measures (including the prevention of death, myocardial infarction, stroke, venous thrombosis and pulmonary embolism) in different categories of patients.[16–18] All three reviews were rigorous in their methodology. The reviews all presented adjusted indirect comparisons of aspirin plus dipyridamole and aspirin alone, using a common control group. The results for the indirect comparisons were presented separately from those of the direct comparison, and used to enhance the findings of the reviews. Conclusions were drawn cautiously in each review.

A fourth review comparing aspirin plus dipyridamole with aspirin alone was conducted by Lowenthal and Buyse, examining the effectiveness of the drugs for the secondary prevention of cerebrovascular accidents.[19] The outcomes of interest were total mortality, vascular mortality, total strokes, fatal strokes, and a composite end-point consisting of vascular death or non-fatal stroke or non-fatal myocardial infarction ('important vascular events'). Double-blind RCTs comparing aspirin with placebo, aspirin plus dipyridamole with placebo, or aspirin plus dipyridamole with aspirin alone were included in the review. Adjusted indirect comparison was made using the placebo groups. Overall odds ratios for aspirin versus placebo were plotted alongside odds ratios for aspirin plus

dipyridamole versus placebo for all outcomes. The risk reduction for each outcome was consistently better for the combination of drugs when compared with aspirin alone. Chi-squared tests with one degree of freedom were calculated to test whether the differences observed could be ascribed to chance alone, and a statistically significant difference was demonstrated in favour of the combination for three of the five outcomes measured [important vascular events (risk reduction 18% versus 40%; $\chi^2 = 7.30$, $p = 0.007$), all strokes (risk reduction 17% versus 42%; $\chi^2 = 7.15$, $p = 0.007$) and fatal strokes (risk reduction –10% versus 43%; $\chi^2 = 4.60$, $p = 0.03$)]. No such statistically significant differences were noted in the trials comparing the two treatment regimens directly. The authors of the review interpret the results with caution. They state that the results "*suggest* that the combination therapy of aspirin with dipyridamole may be superior to aspirin alone". However, they discuss that the results from the indirect comparisons may reflect differences in selection criteria or other confounding factors, rather than a truly greater treatment effect of combination therapy.

Zhang and Li Wan Po conducted a systematic review to assess the efficacy of paracetamol and its combination with codeine or caffeine in comparison to paracetamol alone.[27] An adjusted indirect comparison was made by estimating the pooled effects of each treatment arm against placebo and then by comparing the two estimates. Second, a direct comparison was made between paracetamol–dextropropoxyphene and paracetamol using head-to-head trials (ignoring the placebo group in the three-armed studies). The results were expressed as the difference in percentage improvement of total pain relief (TOTPAR%) and the sum of pain intensity difference. The proportions of patients obtaining moderate to excellent pain relief relative to placebo and the ratio of patients requiring analgesic remedication were also estimated. The results of the indirect comparison were similar to those of the head-to-head comparisons and demonstrated an enhanced analgesic efficacy when codeine (60 mg) was used in addition to paracetamol (600 mg) (using TOTPAR% as the outcome measure).

The efficacy of paracetamol and its combination with codeine was also assessed by Moore and colleagues.[24] They discussed the results of the indirect comparison (between paracetamol plus codeine and paracetamol alone, using placebo as the common control) and direct comparison separately, making no attempt to combine them. In both cases an increased response rate was noted for the combination therapy, although the indirect comparison gave a greater estimate of effect.

Li Wan Po and Zhang conducted a systematic review of RCTs to evaluate the comparative efficacy and tolerability of paracetamol–dextropropoxyphene combination and paracetamol alone.[20] The main outcome measures used in this review were the sum of difference in pain intensity, the response rate ratio and difference in response rate; and the rate ratio and rate difference of side-effects. The paracetamol–dextropropoxyphene combination was compared with paracetamol alone both directly and indirectly (adjusted indirect comparison), as in their previous review.[27] The results of the indirect comparison were used to support the findings of the direct comparisons. The mean (95% CI) difference in the sum of difference in pain intensity, as illustrated by the direct comparisons, was 7.3% (–0.2 to 14.9%) (fixed effect model), in favour of the combination. The results of the indirect comparisons were consistent with the head-to-head comparisons and the authors concluded that on the evidence of both direct and indirect comparisons "there is little objective evidence to support prescribing a combination of paracetamol and dextropropoxyphene in preference to paracetamol alone in moderate pain such as that after surgery".

The benefits of a longer duration interferon regimen (3 MU three times per week for 12 months) in comparison to the standard 6-month regimen for patients with chronic hepatitis C was evaluated by Poynard and colleagues.[26] They presented results of an adjusted indirect comparison, comparing the pooled odds ratios for the different regimens in comparison to undefined controls. For example, the odds ratio for the sustained alanine transaminase (ALT) versus control at the end of the 12-month regimen was 0.35 (95% CI 0.28 to 0.43). This was shown to be greater than for the 6-month regimen, which produced an OR of 0.21 (95% CI 0.13 to 0.28), although the difference was not statistically significant ($p = 0.06$). Direct comparison of a 12-month (or more) regimen and a 6-month regimen showed a statistically significant duration effect on the sustained response rate at 3 MU, OR 0.16 (95% CI 0.9 to 0.23) ($p < 0.01$) in favour of the 12-month regimen.

Matchar and colleagues did not make any reference to the use of indirect comparisons within

their review of medical treatments for stroke prevention, but the pooled relative risks for warfarin and aspirin versus a control/placebo group were compared, as were different doses of aspirin versus placebo.[22] The findings of the indirect comparisons were similar to, and used to reinforce, those of the direct comparison (although only one head-to-head study of warfarin versus aspirin was included).

Piccinelli and co-workers examined the efficacy of drug treatment in obsessive–compulsive disorders.[28] The review found considerable differences between the results from the indirect comparisons and direct comparisons. The authors recognise that the increase in improvement rate was greater for clomipramine than for selective serotonin reuptake inhibitors (SSRIs) when compared with placebo, but highlight the fact that direct (head-to-head) comparisons showed similar therapeutic efficacy on obsessive–compulsive symptoms. Possible reasons for discrepancies are covered in Chapter 6.

Naive indirect comparisons (see Appendix 3, Table 20)

Three other reviews included both direct and indirect comparisons, but did not use an adjusted indirect comparison.[21,23,25] Marshall and Irvine undertook a systematic review to establish the role of rectal corticosteroids in the management of active distal ulcerative colitis.[21] They included RCTs that assigned patients to two or more treatment groups, with rectal corticosteroids in at least one arm. Pooled response rates for each type of corticosteroid and control therapy were calculated across all trials. The pooled response rates for the conventional rectal corticosteroids, the topically active corticosteroids, aminosalicylates and placebo were compared. Conventional rectal and topically active corticosteroids produced similar response rates for symptoms, endoscopy, histology and remission. The aminosalicylates showed improvements in all response rates. The authors of the review sensibly focused on the results of direct comparisons in their discussions, concluding that rectal 5-ASA (an aminosalicylate) is superior to rectal corticosteroids in the management of distal ulcerative colitis. Although the authors did not draw heavily on the findings from the indirect comparisons, the results of these are fairly supportive of the conclusions.

Pope and colleagues also used a naive indirect comparison technique to investigate the hypertensive effects of non-steroidal anti-inflammatory drugs (NSAIDs) and ranked them by magnitude of change in mean arterial pressure (MAP).[25] They did this by extracting data from each NSAID treatment arm across all trials. Data on possible confounders (including age, trial quality, dietary salt intake, hypertensive or normotensive patients) were recorded and adjusted for in the calculation of the average change in MAP for each NSAID. The results of the indirect comparison are used greatly within the review, but are compared with the results of head-to-head comparisons when available. The findings from the direct and indirect comparisons were not always consistent.

The results from indirect comparisons are not always used cautiously. The comparative tolerability and rate of withdrawal from clinical trials of roxithromycin and erythromicin in patients with lower respiratory tract infections were examined in a systematic review, again using both direct and indirect comparisons.[23] Three RCTs included in the review were head-to-head comparisons of roxithromycin and erythromicin. All other trials (n=22) compared either roxithromycin or erythromicin with another macrolide or other agent commonly used as first line therapy for patients with lower respiratory tract infections. Summary statistics of reported adverse events and withdrawals based on data from arms of all trials (both head-to-head and indirect comparisons) were calculated and used to formulate the review's conclusions. Although the results of the three head-to-head trials are presented, summary statistics for these trials alone are not calculated. The authors do discuss the possibility of potential confounding factors, but conclude that there is no significant difference between groups in terms of clinical efficacy, age, gender, settings, duration of treatment, indications or year of publication.

Reviews using indirect comparisons alone

Of the 23 reviews presenting results of indirect comparisons only, 15 used an adjusted method of comparison,[7–9,11,29–39] and eight performed a naive indirect comparison[6,40–46] (*Table 3*).

Adjusted indirect comparisons (see Appendix 3, Table 21)

Of the 15 studies including an adjusted indirect comparison, only six of these interpreted the results in an appropriate way.[8,29–33] For example, Zalcberg and colleagues undertook a systematic review of RCTs to determine the effect of 5-fluorouracil (5-Fu) dose in the adjuvant therapy

TABLE 3 *Reviews using indirect comparisons alone*

	No. of studies	Appropriate interpretation of results		
		Yes	No	Unclear
Adjusted IC	15	6	3	6
Naive IC	8	–	7	1

of colorectal cancer.[29] The trials included in the review compared a 5-FU-containing regimen with a no-chemotherapy control group. For each study, the observed and expected number of deaths on treatment was calculated and an estimated odds ratio of mortality obtained for combined studies using a fixed effect model. Forest plots were presented for trials comparing a 5-FU-containing regimen (separated into four categories: ≥ 10 g, between 8 and 10 g, <8 g and oral chemotherapy) with no-treatment controls, and also for trials comparing 5-FU and levamisole or another 5-FU regimen with no-treatment controls. The authors were cautious in their interpretation of the results stating that, owing to the fact that findings were based on indirect comparisons, confounding by the type of patient being studied in each trial is a possibility.

A review by the Homocysteine Lowering Trialists' Collaboration also used a no-treatment control group as the common comparator to enable indirect comparisons to be made for different doses of folic acid.[8] The aim of the review was to determine the size of reduction in homocysteine concentrations produced by dietary supplementation with folic acid and with vitamin B_{12} or B_6. Twelve trials were included in the review, with individual data on 1114 patients. A forest plot was used to illustrate the reductions in blood homocysteine concentrations with varying doses of folic acid. The findings suggest that a wide range of doses (0.5–5 mg) is similarly effective. The authors of the review provide detailed implications for future research in this area.

In a comparison of the efficacy of home-administrated low molecular weight heparin (LMWH) in the treatment of deep vein thrombosis (DVT) with that of hospital-administered LMWH, Leizorovicz used an unfractionated heparin (UFH) group as the common comparator.[30] Two subgroups of studies were identified (LMWH administered at home versus UFH and LMWH administered at hospital versus UFH) and the odds ratios for recurrent thromboembolic events, mortality and major haemorrhage were presented in a forest plot. The authors report similar efficacy results for the recurrence of thromboembolic events or death, but acknowledge that this approach reduced the statistical power for each subgroup.

Pignon and co-workers conducted a meta-analysis using individual patient data from RCTs comparing chemotherapy alone with chemotherapy combined with thoracic radiotherapy.[31] They clearly acknowledge that indirect comparisons were used to compare trials of early versus late radiotherapy, and also trials with or without sequential radiotherapy. The comparisons did not reveal any optimal time for treatment. The authors conclude that in order to identify the optimal combination of chemotherapy and radiotherapy, further trials, of direct comparisons, are required.

Similarly, a meta-analysis of individual patient data to evaluate the effect of different cytotoxic chemotherapy regimens on patients with non-small cell lung cancer stated that from the trials included it was not possible to recommend one particular regimen over another, and that "Further randomised trials are needed to determine which regimens are the most effective of the modern chemotherapies studied".[32]

Rossouw examined angiographic trials to assess the overall effects of lipid reduction on angiographic outcomes and clinical events.[33] The trials included in the review compared various interventions (lifestyle changes, resins, statins, combinations of drugs, and surgery) with control groups. The author did not discuss the possibility of bias occurring due to the use of indirect comparisons. However, the review demonstrated no evidence of a class effect, with all classes of intervention appearing to have beneficial effects on both angiographic and cardiovascular outcomes.

A review of antiemetic drugs for the prevention of vomiting following paediatric strabismus surgery was conducted by Tramer and colleagues.[11] The review included 24 RCTs that were placebo controlled or no-treatment controlled, or included an unspecified control group. The drugs examined were droperidol (varying doses), metoclopramide (varying doses), dixyrazine, ondansetron, lignocaine, hyoscine, atropine, lorazepam and propofol. Only three of the drugs (droperidol, metoclopramide and propofol) were used to draw conclusions, owing to insufficient

data for the remaining drugs. Adjusted indirect comparisons were made for these three drugs, with odds ratios and 95% confidence intervals being calculated using a fixed effects model, and NNT also calculated. The findings for droperidol suggested a dose–response relationship (10–75 $\mu g\ kg^{-1}$), with 75 $\mu g\ kg^{-1}$ being the only dose to have an odds ratio greater than one. The authors do interpret the result with caution; however, this caution is mainly due to the small sample sizes being examined and not the potential biases that can occur through indirect comparisons.

The appropriateness of the conclusions is unclear in five other systematic reviews also presenting adjusted indirect comparisons. For example, Koch and colleagues[34] conducted a review of H_2 blockers and misoprostol. The trials included had to have a placebo arm to act as the common comparator. The rate difference for each active drug in comparison to placebo was calculated, as were NNTs. Forest plots were presented. The authors claimed that "gastric ulcer was found to be significantly reduced by misoprostol – both in short-term and long-term NSAID treatment – but not by H_2 blockers". In the discussion, the authors did not explicitly discuss biases that can occur through indirect comparisons, but they did highlight the fact that there were differences in the characteristics of patients studied in the misoprostol and H_2-blocker trials (those included in the misoprostol trials were at higher risk of gastric ulcer).

Holme reviewed the association between total mortality outcome or coronary artery disease (CAD) incidence and the amount of cholesterol reduction in randomised cholesterol-lowering trials that were performed before and after inclusion of statin trials.[35] An adjusted indirect comparison was undertaken comparing diet, statins and hormones. The common comparator was fibrates. Multiple regression analysis was undertaken, adjusted by cholesterol changes and baseline risk of CAD. The statistical analyses were done by weighted multiple linear regression models with a fixed effects variance assumption. Diet versus fibrates gave an odds ratio of 0.975 for total mortality and 1.07 ($p < 0.05$) for CAD incidence. Odds ratios of 1.088 and 1.185 ($p < 0.05$) were obtained for hormones versus fibrates, and statins versus fibrates showed an odds ratio of 0.833 ($p < 0.05$) and 0.875 ($p < 0.05$) for total mortality and CAD incidence, respectively. The authors concluded that "Fibrate trials as a group had the least favourable outcome profiles for CAD and all cause mortality of all other drug trials (except hormones)".

Garg and Yusuf[7] conducted a systematic review evaluating the effect of ACE inhibitors on mortality and morbidity in patients with symptomatic congestive heart failure. Thirty-two placebo-controlled trials of ACE inhibitors, including 7105 patients, were used to make adjusted indirect comparisons of estimates of effect for the different agents evaluated. Trials making head-to-head comparisons of different ACE inhibitors were excluded from the review unless they also included a placebo group. The authors state that "Similar benefits were observed with several different ACE inhibitors, although the data were largely based on enalapril maleate, captopril, ramipril, quinapril hydrochloride, and lisinopril".

A systematic review to compare the effectiveness and safety of oral tramadol with standard analgesics using a meta-analysis of individual patients' data included 3453 postoperative patients.[36] Tramadol (50, 75, 100, 150 and 200 mg), codeine (60 mg), aspirin (650 mg) plus codeine (60 mg) and acetaminophen (650 mg) plus proxyphene (100 mg) were all compared with placebo, and then adjusted indirect comparisons made. Relative risks and NNT were presented with 95% confidence intervals, and illustrated graphically. Again, the authors do not discuss the potential biases associated with indirect comparisons. Head-to-head comparisons were possible in certain cases, although data were not presented.

The sixth study for which it was unclear whether the interpretation of the results was appropriate or not was conducted by Lefering and Neugebauer.[39] They compared studies of low-dose corticosteroid versus control with studies of high-dose corticosteroid versus control. The outcome assessed was mortality. The control groups were unspecified, and there was wide variation in the mortality rates among the control groups (7–69%). Pooled rate differences were calculated for the low-dose and high-dose studies, and the results compared. Low-dose corticosteroids showed an effect of –1.9% (95% CI –20.0 to 16.2%), while high-dose corticosteroid trials had a pooled effect of 3.6% (95% CI –2.5 to 9.8%). The findings were inconclusive and the authors state that "Neither the type of steroid used nor the separation into low-dose or high-dose regimen indicated a remarkable difference between the steroid group and control group." No discussion about the role of indirect comparisons was presented.

Boersma and co-workers examined early thrombolytic treatment in acute myocardial

infarction and used indirect comparisons to examine the relationship between time to treatment, from onset of symptoms, and mortality up to 35 days.[9] Odds ratios were calculated and presented graphically. Both linear and non-linear regression analyses were undertaken. The authors conclude that the beneficial effect of fibrinolytic therapy is substantially higher in patients presenting with 2 hours after symptom onset compared with those presenting later. They did not mention the potential problems of indirect comparisons. However, a previous meta-analysis of the same studies argued that "if patient categories can be arranged in some meaningful order then … may be reasonably reliably informative …".[47] The reviews illustrate the usefulness of indirect comparisons when there is a lack of direct comparison information.

The purpose of the systematic review conducted by Aro[37] was to compare the cardiovascular and all-cause mortality of two oestrogen regimens [polyestradiol phosphate (PEP) alone, or in combination with oral ethinyl estradiol (EE)] and orchidectomy with those of the Finnish male population. The review also compared the age-standardised mortality for all three treatment forms. Only two trials were included in the review, Finnprostate I (PEP plus EE versus orchidectomy) and Finnprostate II (PEP versus orchidectomy). Age-specific person-years at risk were computed for each treatment group at 5-year intervals. The authors concluded that "intramuscular PEP monotherapy is associated with low cardiovascular mortality and with an all-cause and prostatic cancer mortality equal to orchidectomy". The results and conclusions of the review are questioned in an article by Ekbom and Taube,[48] who highlight the fact that the two RCTs were conducted in different years and with different inclusion criteria. Indeed, there were statistically significant differences between patient characteristics in the two trials. Ekbom and Taube suggested that the observed low mortality in PEP alone may be due to "flaws in the methodology".[48]

The objective of the meta-analysis carried out by Poynard and colleagues was to compare lansoprazole with raniditine or famotidine in acute duodenal ulcer, and to compare indirectly lansoprazole with other drugs (omeprazole, nizatidine, cimetidine and sucralfate).[38] Raniditine or famotidine was used as a common comparator. Four-week healing rates (OR, 95% CI) were calculated for each drug in comparison to the common comparator, and ranked according to efficacy. The authors mention an RCT directly comparing lansoprazole versus omeprazole in acute duodenal ulceration; however, this trial is not included in the meta-analysis.

Naive indirect comparisons (see Appendix 3, Table 22)

Eight studies, presenting results of indirect comparisons alone, used the naive approach.[6,40–46] None of the studies presented results from direct comparisons (even if this were possible), and potential biases associated with indirect comparisons were rarely discussed.

Coulter and colleagues[41] reviewed RCTs of drugs used to treat menorrhagia. A variety of NSAIDs, antifibrinolytics, hormones and intrauterine devices was examined. The drugs were listed according to the percentage reduction in menstrual blood loss. The review did not present the data from the head-to-head comparisons made within some of the RCTs. Half of the trials included a placebo group and could have been used to carry out an adjusted indirect comparison, but the results of the placebo controls were not reported.

The comparative efficacy and toxicity of second line drugs in rheumatoid arthritis was examined by Felson and colleagues.[6] For each trial, the treatment arms of interest were identified and data extracted. To compare the drugs they used analysis of variance, weighted by treated group size, multiplied by study quality and adjusted for those covariates that had a significant association with each outcome (tender joint count, erythrocyte sedimentation rate and grip strength). For each outcome auranofin was shown to be weaker than the other second line drugs (methotrexate, injectable gold, D-penicillamine, sulfasalazine and antimalarial drugs). It is unclear whether the adjustment for covariates actually reduced or increased biases. Within-study comparisons were ignored, as were studies comparing drugs directly.

The review by Felson and colleagues[6] was quoted by Imperiale and Speroff[45] in support of using naive indirect comparisons of RCT data. They undertook a meta-analysis to examine the efficacy of thromboprophylaxis following total hip replacement. The naive indirect comparison used in the review was based on the premise that the treatment groups were clinically homogeneous in composition. The authors give the impression that their conclusions were based on evidence from RCTs, even though only between-study comparisons were made.

Bansal and Beto[42] compared the efficacy of therapeutic agents used in the treatment of lupus nephritis using outcomes of end-stage renal disease (ESRD) and total mortality, using the same method of indirect comparison as Imperiale and Speroff.[45] They undertook a simple indirect comparison of results of different arms across the studies included in the review. The appropriateness of pooling the data was examined by two methods (*z*-score and heterogeneity test), but the results of the tests were not presented. The review included RCTs and quasi-RCTs, but the results of the studies were used to make between-study comparisons only, therefore losing the power and rigour of the randomisation process. Again, the authors do not mention the potential problems associated with such indirect comparisons, and no evidence from direct comparisons is provided.

Similar problems occur in a review of antihypertensive agents to reduce left ventricular hypertrophy.[43] The study had strict inclusion criteria in that only double-blind RCTs were included. However, analysis was undertaken by combining all treatment arms of the same drug class and weighting them according to the number of patients in each individual study, thus breaking the randomisation procedure. The authors do discuss the importance of randomisation and the validity of studies, although this is not taken into account in the analysis. Within-study comparisons were ignored and no attempt was made to carry out an adjusted indirect comparison using a common placebo group.

Unge and Berstad[44] studied anti-*Helicobacter pylori* regimens. They included both RCTs and observational data, and pooled data into groups according to the combination of drugs used, regardless of, for example, dosage or duration. The authors argued that "a formal meta-analysis is of limited value due to the substantial variation of therapeutic options". There was no discussion of potential biases, and direct comparisons were not included in the review.

Chiba and colleagues[40] conducted a meta-analysis to evaluate the speed of healing and symptom relief in grade II–IV gastrooesophageal reflux disease (GORD). The review included only randomised trials, but did not directly compare different interventions. Data from the included studies were grouped by drug class, decided a priori to be placebo, proton pump inhibitors (PPIs) and H_2-receptor antagonists (H_2RAs). For each study arm the overall healing proportion reported at the final evaluation time-point was used to calculate the overall healing proportion to 12 weeks. Data were pooled within each drug class irrespective of dose, duration of treatment or specific drug. Groups were then compared using analysis of variance (no further details given). Within-study comparisons were not made, although sufficient data to make direct comparisons were presented (see Chapter 6 for further details).

A review comparing the antihypertensive efficacy of available drugs in the angiotensin II antagonist (AIIA) class included 43 trials.[46] Most of the trials were placebo controlled, although some were head-to-head trials. Analysis was based on treatment arms, regardless of which other treatments were included in the trials. The estimate of blood pressure reduction assessed within the review is likely to be overestimated because of regression to the mean effect, as the placebo group (or other comparators) was ignored. The review did not include any recognition of the weakness of this approach.

Summary

Indirect comparisons are commonly used for evaluating the relative effectiveness of alternative interventions. Approximately 9.5% (31/327) of meta-analyses of RCTs identified through DARE included some form of indirect comparison. A further five reviews including some form of indirect comparison were identified. The majority of the reviews included in this chapter were published before 1998, owing to the nature of the searching. An update search was not undertaken as it was felt the sample of reviews was sufficiently large. In addition, there was no reason to suppose that the frequency with which indirect comparisons are used in meta-analyses has altered.

The methods used for the indirect comparisons included the naive indirect comparison (11/36, 31%) and the adjusted indirect comparison (25/36, 69%). Although the identified meta-analyses often included only randomised trials, the strength of the randomisation procedure is completely broken when making naive or unadjusted indirect comparisons, thus providing data that are perhaps equivalent or even inferior to those obtained through non-randomised studies. Such indirect comparisons may be subject to bias (especially selection bias) compared with head-to-head randomised comparisons as the benefit of randomisation does not hold across trials. In 50%

(18/36) of the systematic reviews examined there was no mention of the potential biases associated with the findings of the indirect comparisons.

Results obtained through indirect comparison were not always consistent with the findings obtained by direct comparisons. Thirteen (36%) of the identified systematic reviews used both direct and indirect comparisons. The results of direct and indirect comparisons within these 13 reviews were different in three meta-analyses and similar (same direction but not necessarily magnitude of effect) in eight meta-analyses. Because data from the same trials have often been used in both direct and indirect comparisons in a review, the difference between the direct and indirect estimates may have been underestimated in the reviews.

The indirect comparisons were sometimes carried out implicitly and the results of indirect comparisons interpreted as if from direct comparisons within randomised trials. Further, the findings of direct comparisons were sometimes ignored, even when data were available. The misuse of indirect methods and inappropriate interpretation of results of indirect comparison may result in misleading assessments of relative efficacy of competing healthcare interventions.

Chapter 4

Statistical methods for indirect comparisons

Systematic review of the literature

A systematic review of the research literature was undertaken to identify and evaluate statistical approaches to the analysis of indirect comparisons. The CRD guidelines were used as a starting point for the systematic review protocol.[49] However, as the review was a methodological review rather than a review of the effectiveness of an intervention, it was noted that these guidelines would not be strictly applicable and would need adapting. In particular, the review was not restricted to consideration of publications identified by the formal searching.

Search strategy

A thorough search was undertaken, including both computerised and manual searching, to identify relevant literature. It was recognised that relevant material may also be published in textbooks. Constructing a literature search strategy for methodological papers is problematic owing to the lack of suitable indexing terms in the electronic databases. The development of any search strategy is essentially an iterative process whereby the initial strategy is refined and developed according to its recall and precision. The development of the search strategy in this instance involved five rounds of searching MEDLINE, reviewing the results and adapting the search strategy. The MEDLINE search was run from 1966 to March 1999. It was updated to include records published by February 2001. Details of this process are given in Appendix 5.

Search strategy 5 (Appendix 5) was used as a basis from which to develop strategies to use in other databases. Amendments were made regarding thesaurus terms and subject indexing where appropriate. In general, the strategies used relied heavily on the use of free text terms because of a lack of adequate MeSH or other thesaurus terms to describe the concepts of research methodology. The strategies for PsycLIT (1887 to February 2001), EMBASE (1980 to April 1999), ERIC (1966 to February 1999) and MathSCI (1940 to September 1999) are listed in Appendix 6.

The project team considered that it might be useful to carry out searches of some of the databases covering the agricultural literature. Potential databases to search were identified as being BIOSIS, Agricola, Agris International and CAB Abstracts. Some test search strategies were run on the BIOSIS database, but the records retrieved were studies reporting the results of systematic reviews rather than articles discussing methodological issues. It was decided not to pursue this source further.

In addition to the electronic searches, the following key journals were initially handsearched:

- *Statistics in Medicine* (1984–1999)
- *Controlled Clinical Trials* (1984–1999)
- *Journal of Clinical Epidemiology* (1991–1999)
- *Psychological Bulletin* (1995–1999)
- *Psychological Methods* (1995–1999).

The searches of *Statistics in Medicine*, *Controlled Clinical Trials* and the *Journal of Clinical Epidemiology* were subsequently updated to July 2004. The reference lists of all relevant articles were examined to identify further studies and attempts made to uncover grey and unpublished literature. Contact was also made with those working in the field, both nationally and internationally, including the Cochrane Methods Groups for Empirical Methodological Studies (now the Cochrane Methodology Review Group), Statistics, and Individual Patient Data Meta-analyses. Papers identified ad hoc were also included.

Finally, given the difficulty in identifying relevant articles through electronic searching, an update search was conducted in the form of a citation search run on all of the relevant papers previously identified (April 2004).

Inclusion criteria

The searches were used to identify all papers that addressed the following issues:

- methodology for carrying out indirect comparisons
- methodology for identifying and assessing biases that arise from indirect comparisons
- methodology for avoiding (or even adjusting for) biases arising from indirect comparisons.

To provide examples for the empirical work (see Chapter 5), any reviews meeting the following criteria were also identified:

- systematic reviews reporting both direct and indirect evidence
- examples of systematic reviews where treatment arms from different RCTs are compared (i.e. randomisation has been broken).

Assessment of relevance

All titles and abstracts of articles identified through the searches were screened independently by two reviewers. All papers considered to be relevant by at least one reviewer were retrieved for further appraisal. Retrieved articles were again assessed independently by two reviewers and any disagreements taken to a third party.

Results of searches

In total, 3034 titles and abstracts were identified through the electronic searches. Following screening 29 full text articles were obtained, and further screening resulted in the identification of only six papers for inclusion.[14,50–54] The majority of papers included in this chapter, including two book chapters,[55,56] were identified on an ad hoc basis.

As noted below, the problem of indirect comparisons is closely related to other problems for which articles were not specifically being sought: meta-regression, subgroup analysis, and active–control equivalence trials. Although the search certainly did not detect all relevant publications, it is probable that all of the main methods of analysis were covered.

Statistical methods for indirect comparisons

Background

As noted in Chapter 1, an indirect comparison involves the comparison of the results from sets of studies making different treatment comparisons. It is thus a combination of two (or more) meta-analyses and thus shares all the methodological difficulties associated with the use of meta-analysis to combine the data from several studies. Multiple studies may vary in numerous ways including, but not restricted to, the precise interventions being compared, characteristics of participants, methodological quality (including aspects of treatment allocation and blinding), concomitant interventions, length of follow-up, outcome measures and amount of loss to follow-up. In addition to concerns about the comparability of trials making the same comparison, the comparability of different sets of trials must also be considered. Two further issues are the choice of summary outcome measure[57] and the heterogeneity of the results of the trials, issues that may be related.

The statistical methods for carrying out an indirect comparison can be derived from methods for investigating heterogeneity in a meta-analysis. In meta-analysis it is common to examine separately subgroups of trials, for example, defined by the clinical or demographic characteristics of participants. Such analysis may serve as a means of exploring and possibly explaining statistical heterogeneity of results.[58–60] Subgroups may also be defined by characteristics of one of the treatments. For example, trials may have used two or more different doses of an active treatment, or members of a class of drugs, or may have compared a single treatment against different types of standard or inert treatment. Although it is common to perform separate meta-analyses for each subgroup, a formal approach requires a comparison of the treatment effects in each subset of trials, which assesses the treatment by subgroup interaction. When subgroups define different treatments the analysis is exactly the same as an indirect comparison. Comparison of two independent estimates is a standard statistical analysis, yielding an estimate of the comparative effect in the subgroups, with a confidence interval, and a p-value. The calculation is slightly more complex when the analyses have been of relative measures (odds ratio, relative risk) and thus on the log scale.[61] Alternatively, a regression model can be used to examine whether heterogeneity is explained by one or more study characteristics, known as meta-regression.[60,62] The adjusted indirect comparison can thus be seen to be the simplest form of meta-regression, with a single binary trial factor.

Indirect comparisons are inherent in the use of active–control equivalence drug trials, in which the aim is to demonstrate that a new active drug is equivalent to (i.e. not very different in efficacy from) an already available active drug, which itself has been shown to be superior to placebo.[63,64]

Although indirect comparisons can arise in these different contexts, the statistical analysis options are the same whichever scenario applies. The initial focus is on the simplest, and common, case in which results are available from one or more RCTs comparing A versus B (AvB) and one or

more RCTs comparing AvC, and the aim is to estimate the difference in effect of BvC. The following five sections consider different statistical approaches that have been suggested for indirect comparisons. First, statistical methods using aggregate data are discussed for each study, mainly simple two-step methods. Second, modelling approaches based on individual patient data (IPD) are considered. Although full IPD are rarely available, in the case of binary outcomes the frequencies in a 2×2 table are effectively individual observations and thus such studies are amenable to a wide range of possible analysis methods. In these first two sections the emphasis is on classical frequentist methods that are in wide use and do not require specialist software. Third, Bayesian and likelihood-based methods are considered. Fourth, the assumptions that underlie all indirect comparisons are explored. Finally, various extensions of the approach, including the combination of direct and indirect evidence in a single analysis, are discussed.

The deeply flawed 'naive method', in which the randomisation is broken (described in Chapter 1) is not considered in this chapter, although its performance is evaluated in Appendix 7 and in Chapter 5.

Classical methods using aggregate data

Meta-analysis using summarised (or aggregate) data extracted from published studies is a two-stage process involving the extraction or calculation of an appropriate summary statistic for each of a set of studies, followed by the weighted combination of these statistics to provide an overall estimate of effect (i.e. the contrast between two treatments). Many familiar methods of meta-analysis are of this type, including Mantel–Haenszel methods for binary data and the generic inverse variance method, used for any type of data.[60]

For an indirect comparison these ideas are effectively extended to a three-step approach, in which the third step is to combine the results of two separate meta-analyses into an overall comparison. A standard statistical result is that the variance of the difference between two independent estimates is the sum of the two variances (the variance is the square of the standard error).[65,66] This relation underlies the two-sample *t*-test, for example. It also applies to an indirect comparison, as the two sets of data are from different studies. Thus, given two estimated effects θ_{AB} and θ_{AC} for comparisons of AvB and AvC, respectively, the effect for the comparison BvC is estimated as $\theta_{BC} = \theta_{AB} - \theta_{AC}$, and $\text{var}(\theta_{BC}) = \text{var}(\theta_{AB}) + \text{var}(\theta_{AC})$. A 95% confidence interval for θ_{BC} is obtained as $\theta_{BC} \pm 1.96\sqrt{[\text{var}(\theta_{BC})]}$. The estimates of effect, denoted θ, relate to the scale on which the data would be analysed; examples are risk difference, log risk ratio and log odds ratio for binary data, means for continuous data, and log hazard ratio for time-to-event data. The method in which two separate meta-analyses are combined is referred to as an adjusted indirect comparison. The adjustment here is for a common comparison group. The possibility of further adjustment for covariates is discussed later.

Because the basic method relies on a standard statistical result, it is not possible to say who first applied it in the context of indirect comparison. The earliest methodological discussion found here is by Eddy and colleagues in 1992,[67] but earlier applied examples are likely to exist, especially in the framework of comparing subgroups of trials in a meta-analysis.

The first methodological article explicitly discussing adjusted indirect comparisons seems to be that of Bucher and co-workers.[14] For trials with a binary outcome they suggested combining odds ratios from separate meta-analyses, OR_{AB} and OR_{AC}, so that $\log OR_{BC}$ is estimated as $\log OR_{AB} - \log OR_{AC}$, and its variance as $\text{var}(\log OR_{BC}) = \text{var}(\log OR_{AB}) + \text{var}(\log OR_{AC})$. From these calculations it is simple to obtain a confidence interval for $\log OR_{BC}$ and hence, by transformation, an estimate of OR_{BC} with a confidence interval. The adjusted indirect comparison method is quite general, and this formulation is clearly a specific example of the general method described above. The approach has been used to combine trials using other effect measures, such as risk ratios,[68] risk differences,[68] hazard ratios or means.[69]

Fisher and colleagues discussed the use of adjusted indirect comparisons to address the specific question of estimating the superiority of a drug to placebo when no placebo-controlled trials have been done.[70] This situation arises when one drug has been shown to be effective and it becomes unethical to conduct further placebo-controlled trials of new agents. Baker and Kramer[71] present the same idea from a conceptual perspective. All of their examples are hypothetical and they do not discuss the specifics of analysis.

In a meta-regression analysis, the estimated treatment effect is modelled as a function of one or more study characteristics as predictor variables.[59,62] Least squares regression or a

maximum likelihood method is used and each effect estimate is weighted by the reciprocal of its variance. For a single binary study factor, this analysis is exactly the same as an adjusted indirect comparison. The method can be used to compare results from trials comparing AvB and AvC where the study characteristic is the choice of comparator. The estimated effect comparing BvC is the coefficient for the indicator variable denoting which comparison was made.

All of the authors cited used a fixed effect meta-analysis to combine the results of trials of the same comparison, but the same method can be applied to combine estimates from two random effects meta-analyses. Meta-regression can also be performed allowing estimation of a random effect to describe residual heterogeneity. Only Fisher and co-workers discussed this possibility, and they gave various reasons for preferring a fixed effect meta-analysis.[70] In all meta-analyses a choice has to be made between a fixed effect or random effects analysis; it is not specific to indirect comparisons or meta-regression. However, in this more complex situation there is more than one way to implement a random effects analysis. In a random effects analysis the different trials are assumed to be estimating different, but related quantities that are distributed around some typical value with a variance that is estimated from the data (τ^2). In an indirect comparison, for some models τ^2 is set the same for each component comparison of two treatments, while for other models it is estimated separately.

Methods using generalised linear (mixed) models

The previous section described the analysis of an adjusted indirect comparison in terms of weighted combinations of meta-analyses of sets of trials making different treatment comparisons. An indirect comparison can also be analysed in a regression framework using generalised linear models, but this approach requires the availability of IPD.

For a binary outcome, the 2×2 frequency table for a trial effectively reflects IPD and so such data can be analysed using logistic regression, as discussed by Hasselblad.[52]

The same principles apply to meta-analyses of studies with continuous or time-to-event outcomes, but IPD will need to be obtained from the authors, which is generally a major undertaking.[72] Statistical methods for these outcomes have been considered by Higgins and colleagues[73] and Tudur Smith and colleagues.[74]

For binary and other outcomes, random effects analysis requires fitting generalised linear mixed models (GLMMs),[75] which can be done for example in Stata (Release 8.0; Stata Corporation, Texas, USA, 2003) using the program gllamm, or in SAS (version 9.1) using PROC MIXED (for a continuous outcome) or PROC NLMIXED (for a binary outcome).

These general methods can be used to handle both simple indirect comparisons and more complex networks of treatment comparisons, as discussed below (section 'Extensions to more complex situation', p. 22).

Bayesian and likelihood-based methods

Some authors have presented flexible Bayesian methods that can be used to analyse indirect comparisons as well as more complex data structures. Higgins and Whitehead[53] used a Bayesian approach to investigate the relative effectiveness of β-blockers and sclerotherapy to prevent bleeding in patients with cirrhosis. Twenty-four trials had compared one of the active treatments against control, and two trials had three arms.

The confidence profile method[67] is a very general method for combining almost any evidence relating to a particular question. As well as incorporating studies making different treatment comparisons it can encompass different designs, outcomes and measures of effect, and allows explicit modelling of biases. Although often considered as a fully Bayesian model, it can also be formulated without prior distributions and fitted using maximum likelihood.

Ades[76] recently proposed a Bayesian Markov chain Monte Carlo method to handle the general case. His method of multiparameter evidence synthesis is also very general in allowing other types of evidence to be included in the model. Earlier, Dominici and colleagues[51] presented a hierarchical Bayes grouped random effects model. They used Markov chain Monte Carlo simulation to apply this approach to a highly complex set of 46 trials evaluating treatments for migraine from three classes: β-blockers, calcium channel blockers and biofeedback therapy.

Sutton and colleagues[77] give an overview of the advantages and disadvantages of Bayesian methods in meta-analysis. One advantage of fully Bayesian methods is that they place a distribution on the heterogeneity τ^2 and do not assume that the estimate of τ^2 is without error.[78] This issue is

especially relevant when there are rather few trials making a particular comparison.

Fully Bayesian models offer the greatest flexibility, particularly in modelling random effects. They require specialist software and a deep statistical understanding, taking them beyond the scope of many research groups. Attention in this study is focused on simpler approaches, which can be used for many of the more common scenarios, but the reviewers indicate later where complex methods are needed.

Assumptions

Many study and patient factors (case-mix) influence the outcome of patients in a specific treatment group in a particular trial. The essence of the indirect comparison is that the effect for the comparison BvC is estimated using a common comparator A by contrasting the estimated effects for comparisons of AvB and AvC. Implicit here is the notion that the variation in the observed results for patients treated with the common comparator, treatment A, will (to some extent) account for differences between studies in methodology, case-mix, and so on. The validity of the indirect comparison will thus depend on the extent to which this assumption is reasonable. Adjustment for a common comparator (A) will be expected to reduce, if not remove bias in the comparison of B and C.

The underlying assumption of a fixed effect meta-analysis is that the various studies are all estimating the same effect, such as the effect of A relative to that of B. The same considerations apply to trials comparing AvC. An indirect comparison shares with other observational epidemiological investigations the risk of bias by confounding.[62] The key additional assumption of an indirect comparison using the results of trials of AvB and AvC is that there should be no important differences between the two sets of trials with respect to aspects that could influence (bias) the estimated treatment effect of BvC, that is, there is no confounding of the comparison by some trial characteristic. As an example, Lim and colleagues[79] reported a comparison of low- and medium-dose aspirin therapy after coronary surgery using an indirect comparison of placebo-controlled trials, with outcome evaluated by angiography. There were three randomised trials of low-dose aspirin, with patients receiving angiography at an average of 10, 130 and 180 days, and two trials of medium-dose aspirin, with angiography at an average of 363 and 367 days. Subsequent correspondence[80] considered the possible relation between the time to angiography and the observed effect of aspirin.

Another formulation of this argument is that the two sets of trials should be exchangeable, in the sense that there is no reason to suppose that the results as a whole would be different had the various trialists kept the same protocol and patients, but chosen to study a different treatment comparison.[81] Baker and Kramer[71] show graphically that the validity of the indirect comparison depends on the consistency of the treatment effect over settings with different event rates (different case-mix). Phillips[82] considered some ways in which this assumption might fail. Clearly, one important way in which exchangeability could fail is when the treatment effect is influenced by some factor that itself varies across the different treatment comparisons, such as the clinical setting or length of follow-up. Adjustment for study covariates can help to reduce such an effect and make exchangeability more likely. The analysis is only possible when the factor varies within each set of trials.

Similar issues have been discussed in the context of non-inferiority trials, in which a new treatment is compared with a standard treatment with the aim of showing that the new treatment is no less effective than the standard treatment, within some prespecified margin (these trials are also called active–control equivalence trials).[64,83,84] Often the standard treatment will previously have been shown to be better than placebo. In essence, there is an implicit indirect comparison when inferring from a trial that shows non-inferiority that the new treatment is better than placebo. It is recognised that the validity of such an inference relies on aspects of the non-inferiority trial being closely similar to the prior placebo-controlled trials of the standard treatment. Particular examples include patient characteristics, treatment dose, outcome measures and length of follow-up.[85]

In addition to concerns about the average treatment effect, the issue of heterogeneity has to be considered. Hirotsu and Yamada[54] presented a method that is equivalent to adjusted indirect comparison using inverse variance meta-analysis. They suggested testing for homogeneity of trial results both within treatment comparison and between direct and indirect estimates before pooling.[54] However, an indirect comparison does not require homogeneity, and modelling approaches exist that include random effects terms that estimate the degree of heterogeneity in the comparisons. The model of Higgins and

Whitehead assumes that the degree of heterogeneity is similar in all comparisons,[53] whereas other methods allow separate estimates of heterogeneity for each comparison. Further, in most situations there are too few trials for each paired comparison to allow reliable assessment of whether there is excess heterogeneity, or estimation of separate random effects for each component comparison.

As with any meta-analysis, there is the possibility of fitting random effects models to take account of between-trial heterogeneity. Under a random effects model the treatment effect is allowed to vary among trials making the same treatment comparison (usually assuming that the distribution of effects is normal). Models that apply the random effect to each study arm rather than each comparison between arms are not sensible. In many situations it may be reasonable to assume that the variability of the treatment effects is the same for all paired treatment comparisons.[53]

Several of the papers describing the analysis of indirect comparisons have given rather little consideration to the underlying assumptions, and those statements that have been made are often questionable. For example, the only assumption mentioned by Hasselblad and Kong[68] is that the populations of the individual studies contain some subpopulation in common. While this is sensible, and echoes concerns about equivalence trials, it is by no means a sufficient requirement. Bucher and colleagues[14] noted that "The only requirement is that the magnitude of the treatment effect is constant across differences in the populations' baseline characteristics". In their appendix they mentioned the additional assumption of no treatment by study interaction in each set of trials, but as noted above, the requirement is rather one of equal heterogeneity.

The following section considers various more complex situations involving multiple comparisons of multiple treatments. It is clear that the assumptions become more numerous and tenuous in relation to increasing complexity of the data structure. Most obviously, the notion of exchangeability may need to extend across many sets of trials (which may have been carried out across a long period).

Lastly, in a meta-analysis of trials with a binary outcome, it is well known that the choice of effect measure may have a considerable impact on the analysis, and also on the degree of observed heterogeneity. Empirical studies show that for binary outcomes measures of relative effect (odds ratios and relative risk) are more likely to be consistent across trials than measures of absolute effect (risk difference).[65,86] This issue is of major relevance to indirect comparisons of two or more sets of trials. None of the papers cited above considered the impact of the choice of effect measure in making indirect comparisons, with the odds ratio used in almost all cases. Another assumption of an indirect comparison meta-analysis, therefore, is that the effect measure is appropriate.

Extensions to more complex situations

This section considers six ways in which the available data may be more complex than the simple case outlined above. First, three further situations are considered where all trials still have two treatment arms, and then two cases where at least one trial has more than two arms. Lastly, the case where study or patient characteristics are taken into account is considered.

More than one common comparison treatment (e.g. AvB, AvC, DvB and DvC)

All methods for adjusted indirect comparisons extend simply to the case where there are two or more sets of indirect evidence for the comparison of interest, such as comparing B and C using data from trials comparing AvB and AvC and also trials comparing DvB and DvC. Separate estimates of BvC from adjusted indirect comparisons using two different sets of two-arm trials can be combined using inverse variance weighting. The analysis of such a set of trials is illustrated in Chapter 8.[87]

Lumley[88] suggests comparing the results derived from the different comparators, and suggests a parameter to measure the 'incoherence' of the system, which considers the consistency of a specific estimated contrast between two treatments with the rest of the system (here the estimate from the other route). As with other tests in this context, there may be too few trials of each treatment comparison for this test to have much power.

Combining direct and indirect evidence (e.g. AvB, AvC and BvC)

Often there is both direct and indirect evidence. Traditionally, meta-analysts focus only on direct evidence and do not seek indirect evidence, or they disregard it. Indirect comparisons have mainly been used when there is no direct evidence, or perhaps very little. Thus, there are few examples of indirect and direct evidence being combined. Such a combined analysis is considered here; whether such an analysis is sensible is considered in Chapter 8.

The adjusted indirect comparison method does not explicitly generalise to allow inclusion of direct evidence (RCTs of BvC). It is simple, however, to combine the estimate from the adjusted indirect comparison with data from one or more direct comparisons using inverse variance weighting. The separate meta-analysis estimates need not also be obtained from inverse variance meta-analyses.

Hasselblad showed how logistic regression could be used to analyse a mixture of direct and indirect comparisons.[52] The approach works well for fixed effect models. However, because of the way the model is set up, the random effects model he described allows the outcome per treatment arm to vary randomly across trials rather than the treatment effect itself, so that this approach is not recommended. A proper random effects combined analysis requires the use of mixed models, as discussed above.[75] This issue is also considered in Appendix 7.

Hirotsu and Yamada presented a fixed effect method for estimating odds ratios from a mixture of direct and indirect comparisons, noting that the problem can be seen as an example of an incomplete block design.[54] Although presented rather differently, their method is equivalent to adjusted indirect comparison (possibly combined with direct comparisons) using inverse variance meta-analysis.

A sequence of comparisons (e.g. AvB, BvC, CvD and DvE)

The case of a sequence, or chain, of paired comparisons may arise in the drug development process, particularly where the emphasis is on establishing equivalence of new drugs rather than superiority. If A and B are found to be bioequivalent (often defined as an effect ratio in the range 0.8–1.25), and from separate studies so are B and C and C and D, what can be said about the equivalence of A and D?[63]

Such questions may arise primarily in small bioequivalence studies looking at uptake of drugs via measures such as total and peak exposure,[63] but in principle can also arise in studies of clinical equivalence, and may arise in superiority trials too. An example of such a chain is shown in *Table 18* in Chapter 8.

A simple chain such as AvB, BvC, CvD and DvE can be handled by most of the methods of analysis described. For this example, the contrast AvE can be evaluated via a combination of multiple applications of the simple adjusted indirect comparison.[68]

While such an analysis is clearly easy to carry out, it extends the assumption of exchangeability across three or more sets of trials in the chain, which many may consider to be rather tenuous.

One or more multiarm trials (e.g. AvB, AvC and AvBvC)

An extension of the previous case is when there are multiarm trials comparing all three of the treatments of interest. Although the standard approaches to meta-analysis relate to two-arm trials, trials with three or more treatment arms are surprisingly common, making up about a quarter of parallel group trials in a MEDLINE representative sample.[89] Multiarm trials also cause difficulties in conventional meta-analysis.

The adjusted indirect comparison method cannot take account of trials with three or more arms without either splitting or discarding groups. Multiarm trials can be handled using logistic regression, as this method treats the treatment arm as the unit of data,[52] and by more complex methods.

Gleser and Olkin considered the case where there is a set of RCTs in each of which one or more of several active treatments has been evaluated against a common control.[55] They proposed regression models for obtaining estimates of either the risk difference or (log) odds ratio between each treatment and control, and also for contrasts between the active treatments. Their method is conceptually similar to the adjusted indirect comparison.

When a single multiarm trial is used to estimate two or more paired comparisons those estimates will be correlated when they contain a common arm (e.g. AvB and BvC). This correlation was included in the models presented by Gleser and Olkin[55] and Higgins and Whitehead;[53] it is not taken into consideration in logistic regression.

An arbitrary set of all combinations

The most general case allows for trials comparing various sets of two, three or more of many treatments. (The further possibility of the inclusion of other trial designs, such as cluster or cross-over trials, is not considered here, although in principle these could also be added.) An example is used in Appendix 7 as the basis for illustration of some of the methods of analysis. This scenario goes beyond the main focus of this

report, but such data sets can be analysed using several very general methods, such as that of Eddy and colleagues.[67]

Lumley[88] mentioned the possibility of including multiarm trials, but did not show how his network meta-analysis method extends to that case. The assumptions of a single analysis of a complex set of trials are considerable, and the inner features of the data will be obscured. If such an approach is used it seems desirable also to examine simpler analyses of subsets of treatments and to explore consistency of results.

Adjustment for study and/or patient characteristics

Concerns about the validity of the assumptions may lead to the wish to incorporate study-level covariates in the analysis. These may relate to aspects of the study, such as length of follow-up or whether or not double blind, or study-level summaries of patient characteristics such as mean age. Such adjustment may be felt to remove or reduce meta-confounding and make the studies more likely to conform to the assumption of exchangeability. Adjustment will be more useful for study characteristics as there is very poor power to detect the effect of patient characteristics using study-level summaries.[90,91]

Regression methods, including logistic regression, can be used to adjust for study characteristics. In the rare case where full individual patient data are available, then the whole analysis could be performed using individual patient characteristics, for example using multilevel (hierarchical) modelling.[92]

Summary of available analyses

It has been noted that some complex statistical methods are available to analyse any set of trials. Such methods are inaccessible to many researchers, so it is useful to summarise the types of data that can be analysed using simpler, widely available methods.

The method of adjusted indirect comparison can be extended to any case where all of the trials have just two treatment arms, where necessary by using inverse variance weighted combination of separate indirect estimates. For example, separate indirect and direct estimates of a common treatment effect evidence can be combined. Likewise, indirect and direct estimates can also be combined using the results from separate meta-regression analyses. Adjusted indirect comparison and meta-regression can be used to combine fixed or random effects estimates from subsets of trials, but they cannot handle trials with more than two treatment arms (multiarm trials). However, an advantage of both approaches is that they can be used for any type of outcome measure, including odds ratios, relative risk and risk difference for binary data as well as means or hazard ratios.

For a binary outcome, logistic regression can be used to perform a fixed effect analysis for all of the cases discussed in the previous section, including multiarm trials. Here the odds ratio is the only available effect measure.

When there is at least one multiarm trial, random effects analysis requires hierarchical (mixed) models, and these can also be used for continuous outcomes. Such models appropriately consider the random effects to apply to the treatment effect, in contrast to some of the models discussed above. Their use ideally requires expert statistical assistance.

As noted earlier, several more flexible but complex approaches can be used to handle any mixture of data structures, although not all can deal with any type of outcome measure.

Related literature

The focus of interest here was the combination of results from RCTs, and the authors would not advocate including data from uncontrolled studies. However, methods are available to combine data from RCTs with results of uncontrolled studies. Begg and Pilote[93] proposed a random effects model for combining controlled and uncontrolled studies in the context of incorporating historical controls into a meta-analysis of comparative studies. Their method assumes a distribution for the control group event rate (rather than for the treatment effect, as in conventional random effects meta-analysis) and a fixed treatment effect. The relative weight given to the controlled studies depends on the observed degree of heterogeneity. They illustrated their method using four controlled trials (274 patients) and 12 uncontrolled studies (1708 patients) comparing bone marrow transplantation and chemotherapy for acute non-lymphocytic leukaemia. The method was extended by Li and Begg.[94] Raghunathan[95] presented a conceptually similar approach for combining the results of case–control studies with data from controls in previous studies.

Berkey and colleagues[50] presented a method for combining multiple outcomes in a meta-analysis, to allow a composite analysis when not all

outcomes were assessed (reported) in all trials. Their generalised least squares regression model also extended to analysis of single outcomes from trials comparing multiple treatments, not all of which were studied in each trial. They presented an analysis comparing gold and auranofin for reducing tender joints in patients with rheumatoid arthritis. Of the 44 randomised trials, only nine compared directly the two treatments of interest. Their method does not preserve the link between treatment arms in a given trial and thus cannot be recommended.

Lastly, Büchner and colleagues[96,97] described the prospective design of several trials with a common standard arm, to allow indirect comparison via the common treatment. However, they suggest that in each trial only a small proportion of patients are randomised to the common arm and all the patients across the various trials are used as the common comparator. This method mixes randomised and non-randomised comparisons and runs a serious risk of bias.[98]

Comments

Although indirect comparisons are common, surprisingly few methodological papers have proposed or discussed methods for handling such data. Searching for such papers was hampered by the lack of recognised terminology for indirect comparisons; other terms used in the papers identified include cross-trial (or cross-study) comparison,[82,99] connected comparative experiment,[54] network meta-analysis,[88] mixed comparison,[76] and virtual comparison.[100] Hardly any of the papers identified cited any of the others. It is thus quite possible that other methodological papers were missed, although it is unlikely that there are important omissions regarding methods of analysis.

Some valid approaches were identified for aggregate data that could be applied to simple problems using standard software: the adjusted indirect comparison, meta-regression and, for binary data only, multiple logistic regression (but only fixed effect models).

A particular advantage of a regression modelling approach is the possibility to adjust for other variables available for each study. The hope would be that such adjustment may help to explain some of the heterogeneity within and between groups of trials making the same comparisons. Adjustment for aggregated patient-level variables (such as average age) would be expected not to have much effect in comparison with adjustment with the same variables at the individual level.[90,91] All such analyses are open to all of the problems inherent in meta-regression analysis.[62,91] The regression approach may be seen as a simple version of some of the more complex approaches that can be used.

The use of simple adjusted indirect comparisons to combine two-arm trials is being used more often. Some of these methods can be extended to more complex situations. At the extreme, some advanced methods exist that can handle general problems of networks of comparisons, and sometimes also combine different outcomes, study designs and perhaps external information. Whether it is sensible to include such varied information in a single analysis is open to question. It would seem desirable that, if used, such complex analyses are supplemented by simpler analyses.

Chapter 5 reports on empirical investigations into the performance of several methods of making indirect comparisons, using both fixed and random effects, with and without adjustment for study covariates.

Chapter 5

An empirical investigation of the properties of different statistical methods used for performing indirect comparisons

Comparisons of indirect and direct comparisons, while interesting, cannot provide reliable information about the properties of the indirect approaches. Apart from the problem of small numbers of trials, the main difficulty is the fact that the difference in estimated treatment effect between the two sets of trials may be confounded with differences between the studies.

To compare direct and indirect estimates without such confounding an extensive empirical investigation was carried out using data from a large multicentre randomised controlled trial, the International Stroke Trial (IST).[101] The centres in the study were grouped by region to represent separate trials and results from these 'trials' used to generate both direct and indirect estimates of the same treatment comparison.

As all of the centres used exactly the same protocol, including eligibility criteria, the results of the studies can be said to represent a somewhat optimistic case as, in general, further variability will be introduced by variation in protocols across studies.

The International Stroke Trial (IST)

The aim of the trial was to assess the separate and combined effect of aspirin (300 mg daily) and of subcutaneous heparin [5000 IU twice daily (low dose) and 12,500 IU twice daily (medium dose)] administered for 14 days. Between January 1991 and May 1996 19,435 patients with suspected acute ischaemic stroke entering 467 hospitals in 36 countries were assigned centrally to a treatment using minimisation within 48 hours of symptoms onset.

Details of the study methods and interventions have been described elsewhere.[101] In brief, using a factorial design, half of all patients were randomly allocated to receive aspirin and half to 'avoid aspirin'. For each of these two groups, half of the patients were allocated heparin (low or medium dose) and half to 'avoid heparin'. Placebos were not used.

The primary outcomes were death within 14 days and death or dependency at 6 months. At 6 months fewer patients in the aspirin group were dead or dependent [62.2% versus 63.5%, $p = 0.07$; a difference of 13 (SD 7) per 1000]. After adjustment for baseline stroke severity, the benefit from aspirin was statistically significant [14 events prevented per 1000 patients (SD 6), $p = 0.03$]. Heparin was not found to have any effect (event rate 62.9% in groups who did or did not receive heparin). There was no detectable interaction between aspirin and heparin in the main outcomes.

Adaptations of the trial

Outcome and treatment

Death and dependence at 6 months was considered as the main outcome. Aspirin versus heparin (either dose) was taken as the comparison of interest. The four treatment groups were labelled as follows:

- no treatment: A
- aspirin: B
- heparin: C
- aspirin and heparin: D.

Countries

Data from the different countries in the trial were used to represent multiple studies in a meta-analysis. Only countries with more than 100 patients for each arm were considered. As Italy and the UK enrolled a large number of patients both were split into three new 'countries' by region. Thus, data from 16 countries were analysed. As a consequence of omitting some smaller centres, the overall comparison between

aspirin and heparin is slightly different from that obtained in the actual trial.

Sampling scheme

For each comparison, a sample of 100 patients was drawn at random with replacement from each relevant treatment arm in each country. Thus, the samples drawn were all of the same size regardless of the number of patients actually in the trial in that country. This eliminated complications arising from variation in sample size across trials.

For indirect comparisons, random samples of patients receiving aspirin (B), heparin (C) or neither (A) were used to create meta-analyses with k 'trials' comparing AvB and k comparing AvC, where k = 8, 4, 2 or 1. Various methods for indirect comparisons were used to estimate the contrast between aspirin and heparin. Random samples comparing BvC were also taken from the same $2k$ countries. The whole process was repeated 1000 times.

All analyses were done in Stata 6.0 (Stata Corporation, Texas, USA, 2000). Meta-analyses were undertaken using the metan command.[102,103] Random effects meta-regression was fitted using the metareg command.[102,104]

Analyses

The researchers were interested in the comparison of aspirin and heparin (BvC). Situations were examined where the comparison between aspirin and heparin could be estimated using indirect comparisons. The first situation (case 1) is examined in most detail as it is the most likely to occur. However, it was recognised that indirect comparisons can occur in many different ways. Three possibilities are examined in cases 2–4 using the same basic approach as in case 1. Each of these cases also used data from patients in the fourth arm of the IST trial who had received both aspirin and heparin (denoted D).

Case 1: Indirect comparisons by one route (trials of AvB and AvC)

Estimates based on indirect comparison (using k trials of AvB and k trials of AvC) were compared with estimates from direct comparisons from the same $2k$ countries.

The specific methods used to estimate the BvC treatment effect were as follows.

- *Indirect comparison* of BvC using each of the following methods

Method	Short name used in tables
Fixed effect adjusted indirect comparison[14] (ratio of odds ratios for AvB and AvC)	Adjusted indirect (fixed effect)
Random effects adjusted indirect comparison (DerSimonian and Laird)	Adjusted indirect (random effects)
Logistic regression (fixed effect)	Logistic regression
Random effects meta-regression	Meta-regression (random effects)
Naive method (adding numerators and denominators for treatment arms)	Naive

- *Direct comparison* of BvC in the same $2k$ countries used for indirect comparison

Method	Short name used in tables
Inverse variance meta-analysis	Meta-analysis (fixed effect)
Logistic regression	Logistic regression
Random effects meta-analysis (DerSimonian and Laird)	Meta-analysis (random effects)

In addition, logistic regression analyses were performed for both direct and indirect comparisons adjusting for the three covariates gender, age and risk score, both at the individual level and using study-level summaries.

Analyses were performed for k = 8, 4, 2 or 1, but some methods could not be used reliably for small k.

The scenario just described is asymmetric regarding the three treatments, focusing on BvC as the comparison of interest. As all of the data were available, it was also possible to set up the same analyses with, in turn, AvB or AvC as the comparison of interest. The extent to which these three analyses would indicate the same performance of different methods of analysis depends on the similarity of the amount of heterogeneity between studies for the different comparisons.

Additional analyses were performed to examine the effect of adjusting for patient covariates. Logistic regression was used with adjustment for patients' gender, age and symptom score. Analyses were done using either individual patient data

(k = 8, 4, 2 or 1) or trial-level summary statistics (k = 8 or 4).

Case 2: Indirect comparison of BvC by two routes (trials of AvB, AvC and DvB, DvC)

The same basic approach to sampling was used, with k trials of each of the four comparisons. Indirect comparison of BvC was made using logistic regression. Analyses were performed for k = 4, 2 or 1.

Case 3: Indirect comparisons of BvC by two steps (trials of BvD, DvA and AvC)

Here, k trials of each of three comparisons were generated, and indirect comparisons made combining two adjusted indirect comparisons. Thus, trials of BvD and DvA were combined to obtain an indirect estimate of the comparison BvA, and then this estimate was combined with the estimate for AvC to give an estimate of the comparison BvC. Analyses were performed for k = 4, 2 or 1.

Case 4: Indirect comparisons of BvC from multiarm trials (trials of AvBvD and AvCvD)

Here, k trials of each of two three-arm comparisons were generated, and indirect comparisons made using logistic regression. Analyses were performed for k = 8, 4, 2 or 1.

Some theoretical results

Precision of indirect estimates

Suppose that a set of trials has been performed on sets of patients randomly sampled from the same population. The symbol θ was used to indicate the estimated log odds ratio and a subscript in square brackets to indicate the number of trials being combined to obtain this estimate. For one trial, suppose that the estimated treatment effect θ has variance σ^2 (= $[\mathrm{SE}(\theta)]^2$). Then for a meta-analysis of $2k$ trials, all of the same size and assuming a common true treatment effect, an inverse variance meta-analysis would provide an estimated variance of the treatment effect of $[\mathrm{SE}(\theta_{[2k]})]^2 = \sigma^2/2k$. The expected variance from an indirect comparison based on k trials for each comparison is

$$[\mathrm{SE}(\theta_{\mathrm{AB}[k]} - \theta_{\mathrm{AC}[k]})]^2 = \frac{\sigma^2}{k} + \frac{\sigma^2}{k} = \frac{2\sigma^2}{k}.$$

Thus, one directly randomised trial is as precise as an indirect comparison based on four randomised trials of the same size. Put differently, four times as many similarly sized trials are needed for the indirect approach to have the same power as directly randomised comparisons. This relation will be approximately true when θ is estimated from k trials of varying sizes.

The 4:1 ratio depends on certain assumptions, including equal variances for comparisons AvB, AvC and BvC. Using τ^2 to denote the between-trial variance, one requires

$$\sigma^2_{\mathrm{AB}} + \tau^2_{\mathrm{AB}} = \sigma^2_{\mathrm{AC}} + \tau^2_{\mathrm{AC}} = \sigma^2_{\mathrm{BC}} + \tau^2_{\mathrm{BC}}$$

This assumption cannot be tested unless one has data from all three two-way comparisons, and even then there are unlikely to be enough trials to allow reliable estimation of all of these variances.

Fixed effect methods

As with standard meta-analysis, fixed effects methods for indirect comparisons (the method of Bucher and colleagues[14] and logistic regression) will underestimate $\mathrm{SE}^2(\theta)$ if there is excess heterogeneity. For the indirect comparison between B and C,

$$[\mathrm{SE}(\theta_{\mathrm{AB}[k]} - \theta_{\mathrm{AC}[k]})]^2 = \frac{2\sigma^2}{k} + \tau^2_{\mathrm{AB}} + \tau^2_{\mathrm{AC}}$$

rather than $2\sigma^2/k$ given earlier. Heterogeneity in at least one of the sets of trials will lead to the estimated $\mathrm{SE}^2(\theta)$ being too small. Random effects analyses would therefore seem to be a safer option for indirect comparisons.

The naive method

The naive method treats the data as if they came from a single trial and completely ignores between-trial variance. To take the simplest case with no excess heterogeneity, $\mathrm{SE}^2(\theta)$ for a single trial of size n is σ^2, as before. A naive comparison between k arms of treatment A and k arms of treatment B is equivalent to a single trial of size kn, so that $\mathrm{SE}^2(\theta)$ would be estimated as σ^2/k, which is half of the variance from the adjusted indirect comparison by the method of Bucher and colleagues.[14] Thus, in the case with no heterogeneity the naive method will give standard errors that are too small by a factor of $1/\sqrt{2}$; that is, about 30% too small. When there is between-trial heterogeneity, the underestimation will be even greater.

Results of empirical studies

Tables 4–10 summarise results of analyses of 1000 simulated data sets, constructed as described in the previous section.

The content of the columns in *Tables 4–6* is described in the box below. *Tables 7–10* contain the same information apart from the estimated τ^2. *Tables 8–10* also do not show estimated coverage. In the tables the notation $k + k$ is used to denote indirect comparisons based on two sets of k trials, and similarly for three or four sets of trials in *Tables 8* and *9*.

Column	
2	Estimated odds ratio (obtained as exponential of the value in column 3)
3	Estimated log odds ratio (mean of 1000 estimated values), denoted $\overline{\log OR}$
4	Standard error of the estimated log odds ratio (mean of 1000 estimated values), denoted $\overline{SE(\log OR)}$
5	Standard deviation of 1000 estimates log odds ratios (provides a non-parametric estimate of uncertainty of the estimated log odds ratio), denoted SD($\overline{\log OR}$) Comparison with the standard error (previous column) indicates the extent to which the method of analysis (e.g. logistic regression) underestimates the imprecision of the estimated treatment effect
6	Median value of τ^2 from 1000 analyses (for random effects models only). For indirect values, this is the median of the average of the two values of τ^2 for the two component comparisons Positive values indicate heterogeneity above chance variation between studies. Differences between values for different comparisons (e.g. comparing *Tables 4* and *5*) indicate non-constant heterogeneity for different comparisons
7,8	Coverage of the 95% confidence intervals: percentages of 1000 confidence intervals for the log odds ratio that were wholly below or wholly above the value in column 2 For a good method there should be about 2.5% of confidence intervals that lie wholly below or wholly above the true value
9,10	Percentages of 1000 trials for which the comparison of two treatments was statistically significant ($p < 0.05$) in favour of each treatment The percentages for indirect comparisons can be compared with the direct comparisons to show overoptimistic or conservative methods

Case 1: Indirect comparisons by one route (trials of AvB and AvC)

Table 4 summarises direct and indirect comparisons of aspirin and heparin (BvC) for direct comparisons based on 16, 8, 4 and 2 trials and indirect comparisons based on 8 + 8, 4 + 4, 2+2 and 1 + 1 trials. Thus, for example, the direct comparison involving 16 trials of BvC is based on exactly the same amount of data as the indirect comparison derived from eight trials of AvB and 8 of AvC. For the direct comparisons results are shown for an inverse variance meta-analysis, logistic regression and random effects meta-analysis (DerSimonian and Laird). For the indirect comparisons analyses were made using the same methods (the inverse variance meta-analysis approach is that suggested by Bucher and colleagues[14]), and also by the naive method. Random effects analyses with fewer than four trials per treatment comparison were not performed.

Several features of the results can be noted:

- The estimated odds ratio is effectively the same whichever method was used. The sampling procedures used should lead to this result.
- As predicted, the results from indirect comparisons were less precise than those from direct comparisons. As noted above, a fixed effect indirect comparison of $k+k$ trials would be expected to give estimates with twice the variance as a direct comparison based on $2k$ trials (all of the same size). Thus, the standard error of the indirect estimate is expected to be about $\sqrt{2} = 1.41$ larger than that of the direct estimate.
- Similarly, the standard error for the indirect comparison of $k+k$ trials would be expected to have the same precision as the direct estimate from $k/2$ trials. *Table 4* shows that the standard errors of the fixed effect indirect estimates from 8+8 trials are indeed very similar to those from 4 direct trials, and likewise for indirect 4+4 and direct 2.
- The standard error obtained from the fixed effect analysis will be too small if there is between-trial heterogeneity (beyond random variation). Column 6 shows that there was such heterogeneity in all cases except for $k<4$. For both direct and indirect fixed effect estimates (first two rows of each section) the empirical standard deviations of the 1000 estimated log odds ratios (column 5) exceed the average standard errors (column 4). An exception here is the direct comparisons using all 16 trials, for which there is no obvious explanation.
- Equivalent random effects methods seem to over-correct for heterogeneity, giving standard errors that are larger than the empirical standard deviations of estimated log odds ratios.[105]

- The results for the adjusted indirect inverse variance meta-analysis are almost identical to those using logistic regression. Likewise, the results from the random effects methods (DerSimonian and Laird meta-analysis and random effects meta-regression) agree closely for these data.
- The naive method gives standard errors that are much smaller than those from all of the valid methods. Further, the empirical standard deviations using the naive method were very much larger than those using proper approaches. As a consequence, the apparent standard error of the estimates from the naive method is only about 40% of the correct size.
- Fixed effect direct and indirect methods do not give the right coverage: the proportion of occasions when 95% confidence intervals do not include the correct value (i.e. the direct estimate based on all the trials) exceeds 5% as a consequence of the excess heterogeneity. By contrast, for random effects models the coverage is about 5%. Again, only the direct analyses based on 16 trials do not follow this pattern. The coverage of the naive method is awful, with over 40% of confidence intervals not including the correct value.
- The probability of obtaining a significant result (power) in favour of aspirin was similar for indirect and direct comparisons of equivalent strength (e.g. direct 4 and indirect 8+8). Random effects analyses had a rather lower power than fixed effect methods, but also a reduced risk of a significant result in favour of heparin. As the number of studies reduces, as expected, the power falls and the probability of a significant result in favour of heparin increases.
- The naive method not only gave a much inflated probability of a significant result, but also gave a high risk, 12–20%, of a significant result in favour of heparin. About half of the analyses using the naive method were statistically significant in one direction or the other, for all numbers of trials considered.

Tables 5 and *6* show the same information as in *Table 4*, but for analyses of no treatment versus heparin and no treatment versus aspirin, respectively. Although the broad picture is the same as in *Table 4*, some differences arise from variation in the degree of heterogeneity of the different comparisons. In particular, for the direct comparison between no treatment and heparin (*Table 5*) the between-trial heterogeneity was much less than for the other two comparisons. In fact, τ^2 for this comparison was about one-quarter of τ^2 for the other two comparisons, indicating that the spread (standard deviation) of the distribution of estimates across trials was twice as great for aspirin versus heparin and no treatment versus aspirin as for no treatment versus heparin. This disparity leads to the underestimation of standard errors of estimates being greatest for no treatment versus heparin. In most real applications one would not be able to estimate τ^2 for all three two-way comparisons, so any such variability would be hidden.

The effect of adjustment for patient characteristics was also explored. *Table 7* shows for the aspirin versus heparin comparison the results of additional logistic regression analyses in which adjustment was made for three patient variables: age, gender and prognostic score. The score indicates the number of symptoms (deficits) that patients presented at randomisation. Only symptoms that increased the risk at the 5% level of statistical significance in univariate analysis were considered. The score was thus based on the following deficits: face, arm/hand, leg/foot, dysphasia, hemianopia and visuospatial. Separate adjusted analyses were made using individual data and also using trial-level summaries (only for analyses of at least 8 trials). For comparison purposes, the results of unadjusted logistic regression analyses are repeated from *Table 4*. Only fixed effects analyses were used.

Adjustment for individual patient data had a small effect on the estimated log odds ratio in both direct and indirect analyses. There was also a small inflation of the standard error of the log odds ratio. Adjustment for study-level summary data had a rather greater effect on the estimated log odds ratio in both direct and indirect analyses, giving results nearer to no effect (OR=1) than the analyses based on individual data. There was also a further inflation of the standard error of the log odds ratio in comparison to the analyses using individual data. The standard errors for analyses with adjustment for study-level summaries were about 50% larger than the standard errors of unadjusted estimates. In all cases the standard errors were smaller than the empirical standard deviations of estimates.

TABLE 4 *Results of direct and indirect comparisons of aspirin versus heparin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence)*

						Coverage of 95% CI		% Results with $p < 0.05$	
	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	Median τ^2	Below true value (%)	Above true value (%)	Favours aspirin	Favours heparin
Direct 16 trials									
Meta-analysis (fixed effect)	0.88	–0.131	0.075	0.070		2.0	1.9	41.7	0
Logistic regression	0.88	–0.132	0.075	0.071		2.0	2.4	42.5	0
Meta-analysis (random effects)	0.88	–0.130	0.097	0.072	0.058	0.6	0.6	21.7	0
Indirect (8) + (8)[a] trials									
Adjusted indirect (fixed effect)	0.88	–0.131	0.149	0.171		4.8	4.5	17.2	0.4
Logistic regression	0.87	–0.134	0.148	0.171		4.2	4.3	17.3	0.8
Meta-regression (random effects)	0.87	–0.138	0.181	0.171	0.037	1.0	3.5	12.1	0.1
Adjusted indirect (random effects)	0.88	–0.129	0.189	0.171	0.050	1.2	2.2	9.1	0.2
Naive	0.89	–0.115	0.102	0.255		22.3	20.1	38.1	12.1
Direct 8 trials									
Meta-analysis (fixed effect)	0.88	–0.133	0.106	0.123		3.4	3.9	27.7	0.7
Logistic regression	0.88	–0.128	0.105	0.123		5.3	3.8	26.9	0.5
Meta-analysis (random effects)	0.88	–0.131	0.137	0.125	0.048	2.3	2.7	16.2	0.2
Indirect (4) + (4) trials									
Adjusted indirect (fixed effect)	0.88	–0.131	0.211	0.248		5.6	4.1	12.4	1.7
Logistic regression	0.88	–0.124	0.210	0.254		5.1	4.0	14.2	2.1
Meta-regression (random effects)	0.88	–0.125	0.259	0.245	0.027	2.5	2.5	7.5	0.9
Adjusted indirect (random effects)	0.88	–0.131	0.273	0.258	0.048	2.5	2.8	7.2	0.6
Naive	0.88	–0.131	0.145	0.380		23.3	23.2	35.2	13.9
Direct 4 trials									
Meta-analysis (fixed effect)	0.88	–0.126	0.149	0.184		6.1	4.9	18.5	1.0
Logistic regression	0.87	–0.134	0.149	0.189		6.3	6.2	21.3	1.1
Meta-analysis (random effects)	0.88	–0.130	0.194	0.187	0.028	3.2	3.5	12.2	0.8
Indirect (2) + (2) trials									
Adjusted indirect (fixed effect)	0.88	–0.123	0.299	0.356		5.6	5.2	10.4	1.8
Logistic regression	0.88	–0.131	0.298	0.355		4.8	4.7	9.4	2.9
Meta-regression (random effects)	0.88	–0.128	0.379	0.357	0.001	3.1	3.7	6.6	1.7
Naive	0.89	–0.116	0.208	0.545		21.5	22.0	28.6	17.3

continued

TABLE 4 *Results of direct and indirect comparisons of aspirin versus heparin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence) (cont'd)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	Median τ^2	Coverage of 95% CI		% Results with $p < 0.05$	
						Below true value (%)	Above true value (%)	Favours aspirin	Favours heparin
Direct 2 trials									
Meta-analysis (fixed effect)	0.88	–0.130	0.212	0.270		6.4	6.4	14.5	2.6
Logistic regression	0.88	–0.132	0.211	0.273		5.4	6.5	16.2	2.1
Indirect (1)+(1) trials									
Adjusted indirect (fixed effect)	0.87	–0.139	0.426	0.529		6.7	5.2	9.9	4.2
Logistic regression	0.87	–0.135	0.425	0.533		5.6	5.8	8.9	3.5
Naïve	0.89	–0.116	0.302	0.824		24.1	21.0	31.1	19.8

[a] (Number of studies involved in the comparison aspirin versus no treatment) + (Number of studies involved in heparin versus no treatment).

TABLE 5 *Results of direct and indirect comparisons of no treatment versus heparin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence)*

						Coverage of 95% CI		% Results with $p < 0.05$	
	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	$\text{SD}(\overline{\text{logOR}})$	Median τ^2	Below true value (%)	Above true value (%)	Favours no treatment	Favours heparin
Direct 16 trials									
Meta-analysis (fixed effect)	0.95	–0.049	0.075	0.074		1.8	2.3	9.4	0.7
Logistic regression	0.96	–0.045	0.074	0.073		2.4	2.0	9.2	0.3
Meta-analysis (random effects)	0.96	–0.046	0.085	0.073	0.017	1.3	1.3	6.7	0.2
Indirect (8) + (8)[a] trials									
Adjusted indirect (fixed effect)	0.96	–0.040	0.150	0.190		5.2	6.7	8.8	5.9
Logistic regression	0.96	–0.040	0.148	0.194		8.0	5.7	10.4	4.7
Meta-regression (random effects)	0.95	–0.055	0.194	0.196	0.058	3.6	3.2	4.5	2.1
Adjusted indirect (random effects)	0.95	–0.052	0.200	0.187	0.066	2.6	2.1	4.5	0.9
Naive	0.96	–0.036	0.102	0.243		23.7	18.7	25.4	16.9
Direct 8 trials									
Meta-analysis (fixed effect)	0.95	–0.053	0.106	0.111		3.0	3.3	8.5	1.3
Logistic regression	0.96	–0.046	0.105	0.111		3.7	3.3	6.9	1.4
Meta-analysis (random effects)	0.96	–0.046	0.122	0.115	0.011	2.7	2.0	5.4	0.9
Indirect (4)+(4) trials									
Adjusted indirect (fixed effect)	0.94	–0.056	0.212	0.276		5.9	6.35	8.7	4.2
Logistic regression	0.95	–0.049	0.210	0.275		6.1	7.3	9.9	4.4
Meta-regression (random effects)	0.94	–0.058	0.277	0.267	0.045	3.6	3.4	5.0	1.9
Adjusted indirect (random effects)	0.96	–0.044	0.288	0.280	0.064	3.1	2.9	4.6	2.0
Naive	0.95	–0.051	0.145	0.363		21.8	23	27.3	18.1
Direct 4 trials									
Meta-analysis (fixed effect)	0.95	–0.053	0.150	0.162		2.7	3.1	7.3	0.9
Logistic regression	0.96	–0.037	0.149	0.168		5.1	3.0	5.6	2.8
Meta-analysis (random effects)	0.96	–0.037	0.178	0.162	0.001	2.5	1.5	4.5	1.1
Indirect (2) + (2) trials									
Adjusted indirect (fixed effect)	0.94	–0.056	0.301	0.412		7.2	6.5	10.3	4.6
Logistic regression	0.96	–0.041	0.298	0.402		8.5	6.8	8.1	6.6
Meta-regression (random effects)	0.94	–0.067	0.399	0.396	0.021	4.2	4.8	6.1	3.3
Naive	0.96	–0.039	0.208	0.509		21.9	19.2	22.5	18.4

continued

TABLE 5 *Results of direct and indirect comparisons of no treatment versus heparin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence) (cont'd)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	Median τ^2	Coverage of 95% CI		% Results with $p < 0.05$	
						Below true value (%)	Above true value (%)	Favours no treatment	Favours heparin
Direct 2 trials									
Meta-analysis (fixed effect)	0.95	–0.051	0.212	0.240		2.9	4.1	5.7	2.5
Logistic regression	0.96	–0.046	0.210	0.236		3.9	3.9	5.8	2.0
Indirect (1)+(1) trials									
Adjusted indirect (fixed effect)	0.94	–0.066	0.426	0.572		6.5	6.9	8.8	5.8
Logistic regression	0.95	–0.055	0.425	0.577		6.0	8.1	8.9	5.2
Naive	0.96	–0.038	0.302	0.784		22.1	20.9	23.5	20.2

[a] (Number of studies involved in the comparison aspirin versus no treatment) + (Number of studies involved in heparin versus aspirin).

TABLE 6 *Results of direct and indirect comparisons of no treatment versus aspirin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence)*

						Coverage of 95% CI		% Results with $p < 0.05$	
	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	Median τ^2	Below true value (%)	Above true value (%)	Favours no treatment	Favours aspirin
Direct 16 trials									
Meta-analysis (fixed effect)	1.08	0.079	0.075	0.072		2.4	2.2	0.1	16.5
Logistic regression	1.09	0.086	0.074	0.075		2.9	1.8	0.1	21.9
Meta-analysis (random effects)	1.09	0.084	0.099	0.075	0.065	0.3	0.8	0	8.4
Indirect (8)+(8)[a] trials									
Adjusted indirect (fixed effect)	1.08	0.074	0.150	0.163		5.0	3.4	1.3	9.1
Logistic regression	1.09	0.090	0.149	0.170		5.4	3.4	1.3	11.4
Meta-regression (random effects)	1.09	0.085	0.181	0.165	0.037	2.5	1.5	0.4	6.2
Adjusted indirect (random effects)	1.09	0.089	0.186	0.167	0.044	2.4	1.7	0.3	5.8
Naive	1.09	0.093	0.102	0.246		24.3	19.4	11.8	34.5
Direct 8 trials									
Meta-analysis (fixed effect)	1.08	0.078	0.106	0.124		5.2	4.3	1.0	14.3
Logistic regression	1.08	0.080	0.105	0.129		6.1	5.8	1.6	15.0
Meta-analysis (random effects)	1.09	0.085	0.140	0.129	0.059	2.3	4.1	0.2	9.5
Indirect (4)+(4) trials									
Adjusted indirect (fixed effect)	1.08	0.074	0.212	0.236		5.3	4.1	1.8	7.8
Logistic regression	1.08	0.076	0.210	0.243		3.9	4.6	1.9	6.9
Meta-regression (random effects)	1.07	0.072	0.262	0.248	0.027	3.2	2.8	1.4	6.0
Adjusted indirect (random effects)	1.09	0.083	0.269	0.239	0.045	2.0	1.7	0.9	4.5
Naive	1.08	0.077	0.145	0.375		23.7	22.9	16.6	31.0
Direct 4 trials									
Meta-analysis (fixed effect)	1.08	0.077	0.149	0.188		6.1	7.4	2.5	13.5
Logistic regression	1.09	0.088	0.149	0.189		7.0	4.6	2.5	14.9
Meta-analysis (random effects)	1.08	0.077	0.198	0.192	0.039	2.8	4.2	1.3	9.8
Indirect (2)+(2) trials									
Adjusted indirect (fixed effect)	1.10	0.094	0.301	0.353		5.5	5.4	2.8	8.4
Logistic regression	1.09	0.088	0.299	0.355		5.8	3.7	1.8	8.4
Meta-regression (random effects)	1.09	0.088	0.377	0.361	0	4.4	2.8	1.9	6.3
Naive	1.08	0.081	0.207	0.542		22.2	22.0	18.0	27.2

continued

TABLE 6 *Results of direct and indirect comparisons of no treatment versus aspirin using data from different numbers of countries as separate trials (results based on 1000 resamplings based on the IST study; patient outcome was death or dependence) (cont'd)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	Median τ^2	Coverage of 95% CI		% Results with $p < 0.05$	
						Below true value (%)	Above true value (%)	Favours no treatment	Favours aspirin
Direct 2 trials									
Meta-analysis (fixed effect)	1.08	0.082	0.212	0.281		4.7	7.9	3.8	11.6
Logistic regression	1.09	0.085	0.210	0.277		7.4	6.1	4.0	12.2
Indirect (1)+(1) trials									
Adjusted indirect (fixed effect)	1.10	0.092	0.428	0.505		4.2	5.4	3.7	7.3
Logistic regression	1.08	0.080	0.427	0.499		5.1	4.5	3.1	6.2
Naive	1.10	0.094	0.300	0.821		23.5	22.1	20.7	27.5

[a] (Number of studies involved in the comparison aspirin versus heparin) + (Number of studies involved in heparin versus no treatment).

TABLE 7 *Results of direct and indirect comparisons of aspirin versus heparin using logistic regression with and without adjustment for patient factors at study and individual level*

					Coverage of 95% CI		% Results with $p < 0.05$	
	OR	**$\overline{\text{logOR}}$**	**$\overline{\text{SE(logOR)}}$**	**SD($\overline{\text{logOR}}$)**	**Below true value (%)**	**Above true value (%)**	**Favours aspirin**	**Favours heparin**
Direct 16 trials								
Logistic regression	0.88	–0.132	0.075	0.071	2.0	2.4	42.5	0
Adjusted for gender, age, score (study level)	0.92	–0.079	0.089	0.089	4.0	1.1	14.1	0.1
Adjusted for gender, age, score (individual level)	0.89	–0.113	0.082	0.080	1.7	2.4	27.6	0
Indirect (8) + (8)[a] trials								
Logistic regression	0.87	–0.134	0.148	0.171	4.2	4.3	17.3	0.8
Adjusted for gender, age, score (study level)	0.92	–0.087	0.171	0.183	3.8	2.8	9.0	1.0
Adjusted for gender, age, score (individual level)	0.90	–0.108	0.164	0.187	4.9	5.6	12.5	1.1
Direct 8 trials								
Logistic regression	0.88	–0.128	0.105	0.123	5.3	3.8	26.9	0.5
Adjusted for gender, age, score (study level)	0.92	–0.078	0.154	0.181	3.9	2.7	11.9	1.0
Adjusted for gender, age, score (individual level)	0.89	–0.113	0.116	0.132	3.5	5.1	20.0	0.2
Indirect (4) + (4) trials								
Logistic regression	0.88	–0.124	0.210	0.254	5.1	4.0	14.2	2.1
Adjusted for gender, age, score (study level)	0.93	–0.071	0.318	0.376	4.9	2.6	5.3	2.0
Adjusted for gender, age, score (individual level)	0.89	–0.115	0.232	0.267	4.3	4.8	8.8	1.9
Direct 4 trials								
Logistic regression	0.87	–0.134	0.149	0.189	6.3	6.2	21.3	1.1
Adjusted for gender, age, score (individual level)	0.89	–0.116	0.165	0.206	5.2	6.1	16.3	1.4
Indirect (2) + (2) trials								
Logistic regression	0.88	–0.131	0.298	0.355	4.8	4.7	9.4	2.9
Adjusted for gender, age, score (individual level)	0.89	–0.113	0.331	0.397	5.4	5.6	8.7	2.8
Direct 2 trials								
Logistic regression	0.88	–0.132	0.211	0.273	5.4	6.5	16.2	2.1
Adjusted for gender, age, score (individual level)	0.89	–0.120	0.236	0.298	4.7	7.3	13.8	2.1
Indirect (1)+(1) trials								
Logistic regression	0.87	–0.135	0.425	0.533	5.6	5.8	8.9	3.5
Adjusted for gender, age, score (individual level)	0.89	–0.116	0.477	0.556	4.1	4.7	6.4	2.8

[a] (Number of studies involved in the comparison aspirin versus no treatment) + (Number of studies involved in heparin versus no treatment).

TABLE 8 *Results of indirect comparisons of aspirin versus heparin (BvC) by two routes (trials of AvB, AvC, DvB and DvC) (case 2) (In parentheses are shown results from* Table 4 *for the same numbers of patients in simple indirect comparison)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	% Results with $p < 0.05$	
					Favours aspirin	Favours heparin
Logistic regression						
Indirect 4 + 4 + 4 + 4 (indirect 8 + 8)	0.88	–0.131 (–0.134)	0.148 (0.148)	0.176 (0.171)	16.6 (17.3)	0.5 (0.8)
Indirect 2 + 2 + 2 + 2 (indirect 4 + 4)	0.88	–0.127 (–0.124)	0.211 (0.210)	0.259 (0.254)	14.6 (14.2)	1.5 (2.1)
Indirect 1 + 1 + 1 + 1 (indirect 2 + 2)	0.88	–0.129 (–0.131)	0.300 (0.298)	0.375 (0.355)	11.3 (9.4)	2.8 (2.9)
Adjusted indirect (random effects)						
Indirect 4 + 4 + 4 + 4	0.87	–0.134	0.188	0.185	11.3	0.4

TABLE 9 *Results of indirect comparisons of aspirin versus heparin (BvC) by two steps (trials of trials of BvD, DvA and AvC) (case 3)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	% Results with $p < 0.05$	
					Favours aspirin	Favours heparin
Adjusted indirect (fixed effect)						
Indirect 4 + 4 + 4	0.89	–0.118	0.258	0.304	11.4	1.8
Indirect 2 + 2 + 2	0.86	–0.156	0.366	0.447	10.4	2.6
Indirect 1 + 1 + 1	0.89	–0.114	0.522	0.637	8.0	3.3
Adjusted indirect (random effects)						
Indirect 4 + 4 + 4	0.88	–0.123	0.335	0.317	5.7	0.5

TABLE 10 *Results of indirect comparisons of aspirin versus heparin (BvC) from multiarm trials (trials of AvBvD and AvCvD) by logistic regression (case 4) (in parentheses are shown results from* Table 4 *for the simple indirect comparison based on AvB and AvC)*

	OR	$\overline{\text{logOR}}$	$\overline{\text{SE(logOR)}}$	SD($\overline{\text{logOR}}$)	% Results with $p < 0.05$	
					Favours aspirin	Favours heparin
Logistic regression						
Indirect 8+8	0.88	–0.131 (–0.134)	0.128 (0.148)	0.152 (0.171)	21.4 (17.3)	0.6 (0.8)
Indirect 4+4	0.88	–0.131 (–0.124)	0.182 (0.210)	0.225 (0.254)	15.4 (14.2)	1.5 (2.1)
Indirect 2+2	0.88	–0.124 (–0.131)	0.258 (0.298)	0.334 (0.355)	12.8 (9.4)	3.0 (2.9)
Indirect 1+1	0.89	–0.120 (–0.135)	0.369 (0.425)	0.468 (0.533)	9.7 (8.9)	2.8 (3.5)

Case 2: Indirect comparison of BvC by two routes (trials of AvB, AvC and DvB, DvC)

Tables 4–7 consider the simplest indirect comparison, in which the contrast BvC is estimated using results from trials of AvB and AvC. *Table 8* shows results from analyses that included a second pathway for the indirect comparison of BvC, namely trials of DvB and DvC. Results for logistic regressions are shown for 4, 2 and 1 trial per comparison. For comparison the results from *Table 4* are repeated here relating to a single pathway of twice as many trials: each of the first four rows of the table thus compares analyses of the same amount of data. The two sets of results are very similar, suggesting that the two approaches are equivalent. In theory they are, but in practice additional heterogeneity variances are playing a part and in general one might expect that the inclusion of additional links would increase the risk of obtaining an erroneous result.

Table 8 also shows the results of a random effects analysis for the case of 4 + 4 + 4 + 4 trials. As was seen in *Table 4*, the standard error of the estimated log odds ratio is larger for the random effects analysis than for the fixed effect analysis (first row).

Case 3: Indirect comparisons of BvC by two steps (trials of BvD, DvA and AvC)

Table 9 shows results from three sets of trials involving four treatments. In effect, this method involves two applications of the adjusted indirect approach in sequence. These results are broadly compatible with those in the previous tables. Again, a random effects analysis is shown for the largest number of trials (4+4+4).

Case 4: Indirect comparisons of BvC from multiarm trials (trials of AvBvD and AvCvD)

Lastly, *Table 10* shows results from indirect comparisons of two sets of three-arm trials. Although these analyses also make use of treatment D, in this set-up the comparison is internal (and randomised), unlike in case 2. Although there is the same number of trials the number of patients is 50% greater than in *Table 8*. As a result, there is a 25% reduction in variance of the estimated log odds ratio compared with the results in *Table 8*. Similarly, there is a reduction in variance compared with the simple indirect comparison based on AvB and AvC, which for comparison is repeated here from *Table 4*.

Summary

Several methods have been proposed to estimate the different effectiveness of treatments not compared directly in randomised trials. This investigation focused on variations of widely available methods: the adjusted indirect comparison (fixed or random effects), meta-regression and logistic regression, and also evaluated the inappropriate 'naive' method.

The results of the resampling studies show that all of the appropriate methods are unbiased and will give the right answer on average across many such applications. However, the correctness of the estimated standard error will depend on strong but unverifiable assumptions. Little difference was found in the performance of inverse variance methods (adjusted indirect comparison) and logistic regression. But, for this data set, the results support the use of random effects versions of either of these approaches. Such a finding is not surprising given that there are (at least) three heterogeneity variances that could have an impact on the results of even the simplest indirect comparison.

With binary outcomes, logistic regression may seem preferable to the adjusted indirect comparison. The results of the two methods are extremely close when both can be applied. An advantage of the adjusted indirect comparison for binary data is that the meta-analysis is not restricted to the use of the odds ratio, although it may be possible to use other types of regression model. The main advantages of logistic regression are that it can be extended to more complex situations and it can be used to adjust for individual- or study-level covariates. In practice, however, IPD are unlikely to be available and the use of study-level patient covariates (such as mean age) is in general too blunt a method to be informative. Certainly, in this investigation adjustment for study-level summaries of patient characteristics led to biased estimates of treatment effect with inflated standard errors. By contrast, the ability to include study-level covariates that relate to the study itself can be valuable, as in the study by Packer and colleagues.[69]

This resampling study used the odds ratio as a measure of effect. While the odds ratio has desirable statistical properties, it may not always be the preferred effect measure for a meta-analysis.[57] Adjusted indirect comparisons based on contrasting separate meta-analyses, such as the approach of Bucher and colleagues,[14] can be

extended very simply to risk ratio or risk difference, but logistic regression and some more complex methods do require the effect measure to be the odds ratio. For continuous data the adjusted indirect comparison approach can be used for either actual or standardised mean differences, as in the study by Packer and colleagues.[69] The choice here is perhaps simpler, with a strong preference for using the actual mean difference as long as the outcome being measured is the same in the different studies.

These analyses were all based on the data from a single large trial, with a unified protocol and hence consistent inclusion criteria. In general, more variation would be expected between independent component studies contributing to an indirect comparison. In particular, there may be differences between the participants in the trials (and aspects of the trials themselves) between the two sets of trials being combined.

In addition, the present analyses compared a treatment with a small benefit (aspirin) with one with no apparent benefit (heparin). The extent to which our findings would apply in the more common case where indirect comparisons are made when each of two active drugs is better than placebo (or no treatment) is unknown. Also, differences between direct and indirect estimates may be less important where the effect of interest was large.

Chapter 6

Statistical discrepancy between the direct and indirect estimate: empirical evidence from published meta-analyses

The objective of this chapter is to summarise empirical evidence about the validity of indirect comparison by measuring discrepancy between the direct and the indirect estimates in a sample of published meta-analyses. It builds on the findings of the survey of the use of indirect comparison for evaluating relative efficacy of competing interventions presented in Chapter 3. (Note: data in this chapter have previously been published.[106])

Methods

Identification of relevant reviews

Detailed strategies for identifying relevant meta-analyses have been described in Chapter 3. In brief, the hard copies of reviews on DARE (1994 to December 1998) were available and read. In addition, the reviews on The Cochrane Library (Issue 3, 2000) were screened. Included meta-analyses had to meet two criteria: competing interventions could be compared both directly and indirectly, and the same primary studies were not used in both the direct and indirect comparison.

Comparison methods

The relative efficacy of competing interventions was estimated using three comparative methods: the direct (head-to-head) comparison, the adjusted indirect comparison and the naive indirect comparison. For the direct comparisons, the relative efficacy (T_{AB}) of intervention A versus B was estimated by comparing the result of group A and the result of group B within a randomised trial. When there were two or more similar trials that compared the same interventions, results of individual trials were weighted by the inverse of corresponding variances and then quantitatively combined. Whenever possible, the random effects model was used for the quantitative pooling.

The naive indirect estimate of relative efficacy was obtained simply by comparing the result of treatment A in trial 1 with the result of treatment B in trial 2. That is, in a naive indirect comparison, results of individual arms between different trials are compared as if they are from a single trial. The variance in the naive indirect comparison will be similar to the variance in the direct comparison with the same sample size (see Chapter 5). When more than one study was available for a treatment, the results of individual studies were weighted by the number of participants in the corresponding arms and then quantitatively combined (thus, between-trial variability was not considered).

Adjusted indirect comparisons were conducted using the method suggested by Bucher and colleagues.[14] In brief, the indirect comparison of interventions A and B was adjusted by the results of their direct comparisons with a common intervention C. Suppose T_{AC} is the estimate of direct comparison of intervention A versus C, and T_{BC} is the direct comparison of intervention B versus C. Then the estimate of the adjusted indirect comparison of intervention A versus B (T'_{AB}, e.g. log relative risk, mean difference) is estimated by

$$T'_{AB} = T_{AC} - T_{BC}$$

and its variance is

$$\mathrm{Var}(T'_{AB}) = \mathrm{Var}(T_{AC}) + \mathrm{Var}(T_{BC})$$

This adjusted method aims to overcome the potential problem of different prognostic factors between study participants in different trials, and it is valid if the relative efficacy of interventions is consistent in patients across different trials. It should be noted that the variance in the adjusted indirect comparison using four trials is equivalent to the variance in the direct comparison within one trial of the same size (see Chapter 5). Multiple studies used in the adjusted indirect comparison are combined by a random effects model whenever possible.

Measures of discrepancy

The relative efficacy was measured using mean difference for continuous data and log relative risk

for binary data. The discrepancy between the direct estimate (T_{BC}) and the adjusted indirect estimate (T'_{BC}) was measured by the difference (X) between the two estimates:

$$X = T_{BC} - T'_{BC}$$

and its standard error is

$$SE(X) = \sqrt{SE(T_{BC})^2 + SE\ (T'_{BC})^2}$$

where $SE(T_{BC})$ and $SE(T'_{BC})$ are the estimated standard errors for the direct estimate and the adjusted indirect estimate, respectively. The 95% confidence interval for the estimated discrepancy was calculated by $X \pm 1.96 * SE(X)$. The estimated discrepancy can also be standardised by its standard error to obtain a value of $z = X/SE(X)$.

In addition, the results of meta-analyses were categorised as statistically non-significant ($p > 0.05$) or statistically significant ($p \leq 0.05$). The statistically significant effect can be further separated according to whether intervention B was less or more effective than intervention C. The degree of agreement in statistical conclusions between the direct and indirect method was assessed by a weighted kappa.

Results

The searches identified 28 systematic reviews in which both the direct and indirect comparison of competing interventions could be conducted, although indirect comparison was not explicitly used in many of these meta-analyses (see Appendix 8). Some systematic reviews assessed more than two active interventions, and a total of 44 comparisons could be made using data from the 28 systematic reviews.

Figure 2 summarises the statistical discrepancies (z statistics) between the direct and both naive and adjusted indirect estimates for 43 meta-analyses (note that one meta-analysis in which the naive indirect comparison was not available was not included in *Figure 2*). There are several significant discrepancies between the direct and indirect estimates, but the direction of the difference is inconsistent. The relative efficacy of an intervention may be overestimated or underestimated by the indirect comparison, as compared with the results of the direct comparison.

Significant discrepancy ($|z| \geq 1.96$) was observed in three of the 44 comparisons between the direct and adjusted indirect estimate (7%). Between the direct and the naive indirect estimate, 11 of the 43 comparisons showed statistically significant discrepancy (26%). The statistical discrepancies between the direct and the naive indirect estimate are generally greater than those between the direct and the adjusted indirect estimate (*Figure 2*).

Figure 3 shows the relation between the statistical discrepancy (z) and the number of trials used in indirect comparisons. Visually, statistical discrepancies tended to be smaller when the number of trials was large (>60) than when the number of trials was small (<40). However, such tendency was not consistent, as the discrepancy between the direct and indirect estimate may be significant even when more than 40 trials have been used for the indirect comparison (e.g. meta-analysis 44).

Figure 4 summarises the differences (and their 95% confidence intervals) between the direct and the adjusted indirect estimates. Significant discrepancy ($p < 0.05$) was observed in three of the 44 comparisons (i.e. the 95% confidence interval did not include zero). In four other meta-analyses, the discrepancy between the direct and the adjusted indirect estimate was borderline significant (i.e. $p < 0.1$). The relative efficacy of an intervention was equally likely to be overestimated or underestimated by the indirect comparison, as compared with the results of the direct comparison.

There was a moderate agreement in statistical conclusions between the direct and the adjusted indirect method (weighted kappa = 0.53) (*Table 11*). In terms of statistical conclusions, 32 of the 44 indirect estimates fell within the same categories of the direct estimates. According to the direct comparison, 19 of the 44 comparisons suggested a statistically significant difference ($p < 0.05$) between competing interventions. Compared with direct estimates, the adjusted indirect estimates were less likely to be statistically significant. Ten of the 19 significant direct estimates became statistically non-significant in the adjusted indirect comparison, whereas only two of the 25 non-significant direct estimates were significant in the adjusted indirect comparison. The agreement between the direct and the naive indirect comparison is much poorer (weighted kappa = 0.28).

Commentary

A wide range of medical topics has been covered by the 44 meta-analyses used in this study

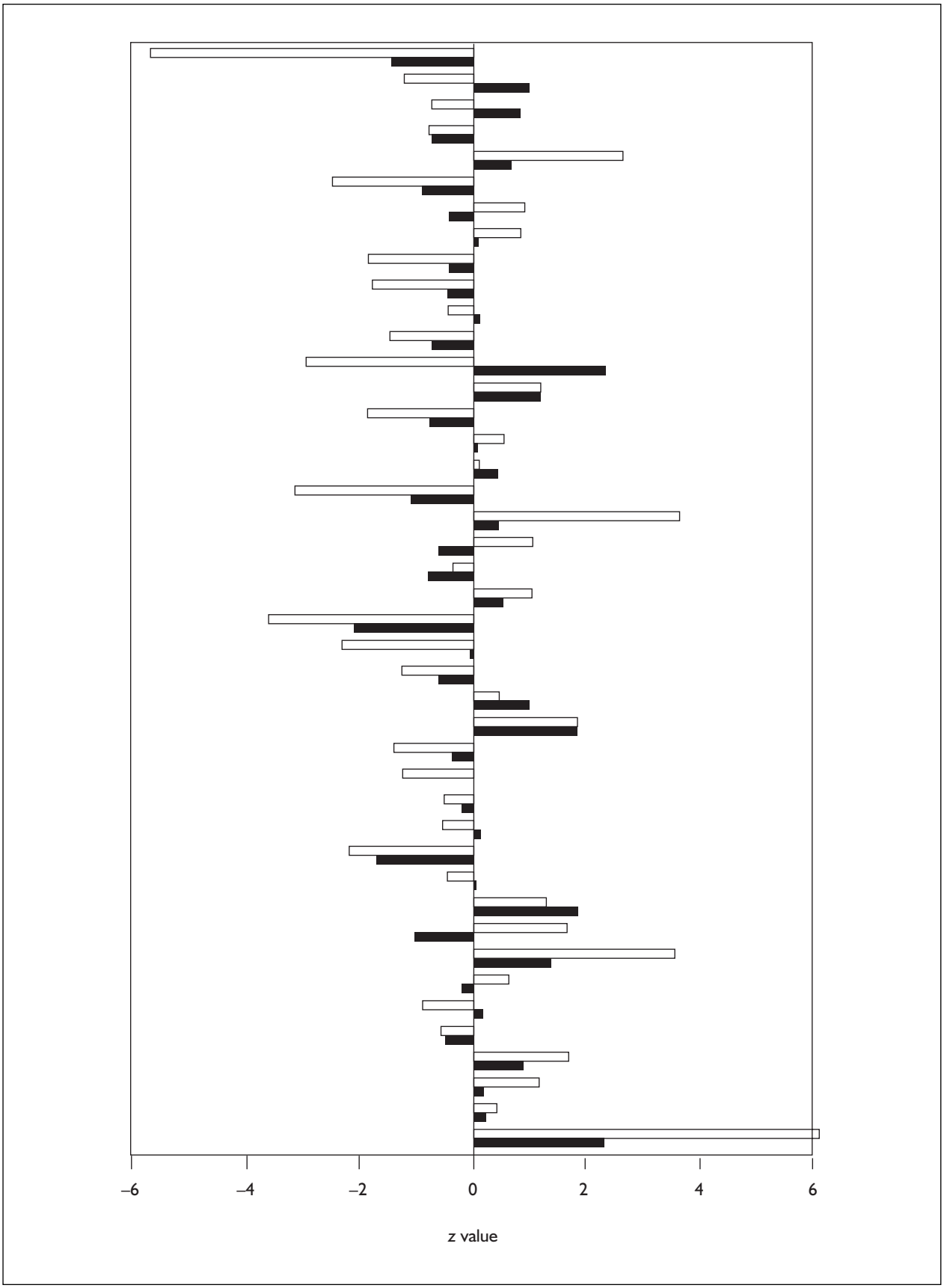

FIGURE 2 *Statistical discrepancy between the direct and the indirect estimate: empirical evidence from published meta-analyses. Solid bars, discrepancy between the direct and adjusted indirect estimate; blank bars, discrepancy between the direct and naive indirect estimate.*

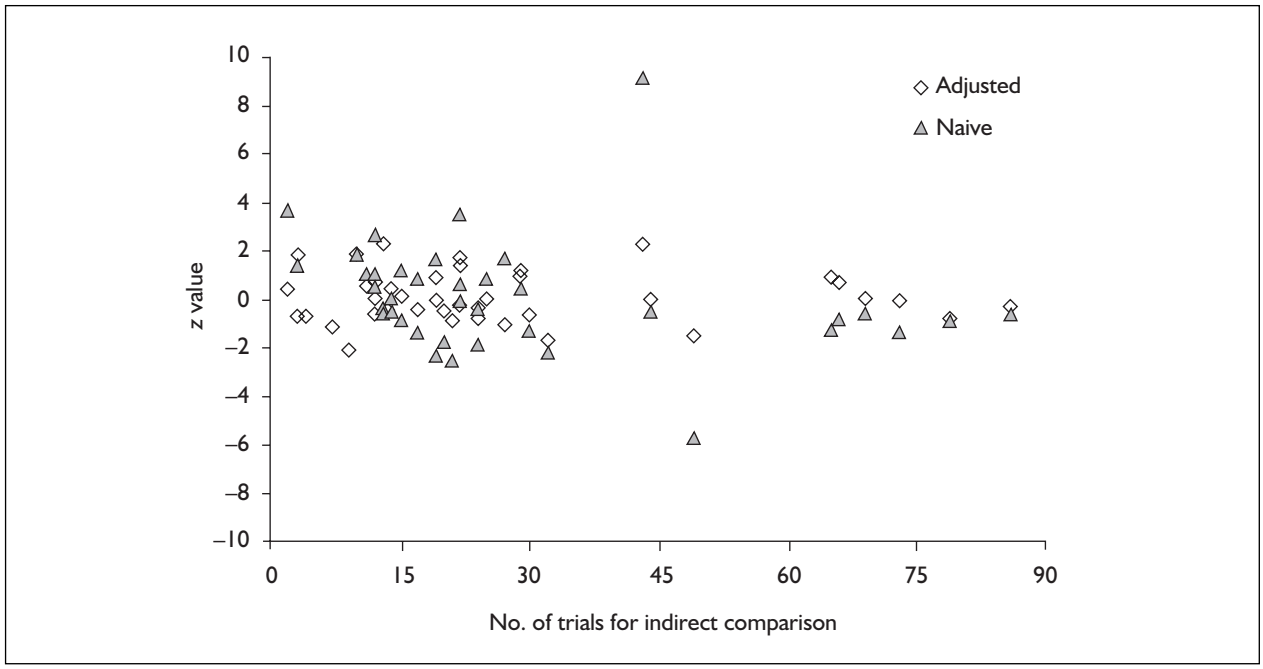

FIGURE 3 *Statistical discrepancy (z value) and the number of trials used in indirect comparison*

TABLE 11 *Methods of comparison and number of significant findings in 44 meta-analyses of competing interventions*

Direct estimate	Adjusted indirect estimate			Naive indirect estimate		
	Significant effect (–) (n = 6)	Non-significant effect (n = 33)	Significant effect (+) (n = 5)	Significant effect (–) (n = 9)	Non-significant effect (n = 19)	Significant effect (+) (n = 15)
Significant effect (–) (n = 8)	5	3	0	3	5	0
Non-significant effect (n = 25 or 24)	1	23	1	5	11	8
Significant effect (+) (n = 11)	0	7	4	1	3	7

Non-significant effect: difference between intervention groups is statistically non-significant ($p > 0.05$); significant effect ($p \leq 0.05$) is separated according to whether the intervention A is less (–) or more effective (+) than intervention B. Agreement between the direct and the adjusted indirect estimate: weighted kappa value 0.53; agreement between the direct and the naive indirect estimate: weighted kappa value 0.28.

(Appendix 8). The patient categories include those with an increased risk of vascular occlusion, HIV-infected patients, those with GORD, postoperative pain or dyspepsia, and cigarette smokers. These meta-analyses were used to obtain data for examining discrepancies between the different comparative methods. The study did not critically appraise the methodological quality of these meta-analyses and of primary trials in these meta-analyses.

The findings presented in this paper suggest that the direction of discrepancy between the direct and the indirect estimate is unpredictable, but the discrepancy in the adjusted indirect comparison is generally less than that in the naive indirect comparison. The observed significant discrepancies between the direct and the indirect estimates may be explained by many possible factors, such as random errors, baseline prognostic characteristics, interventions other than those

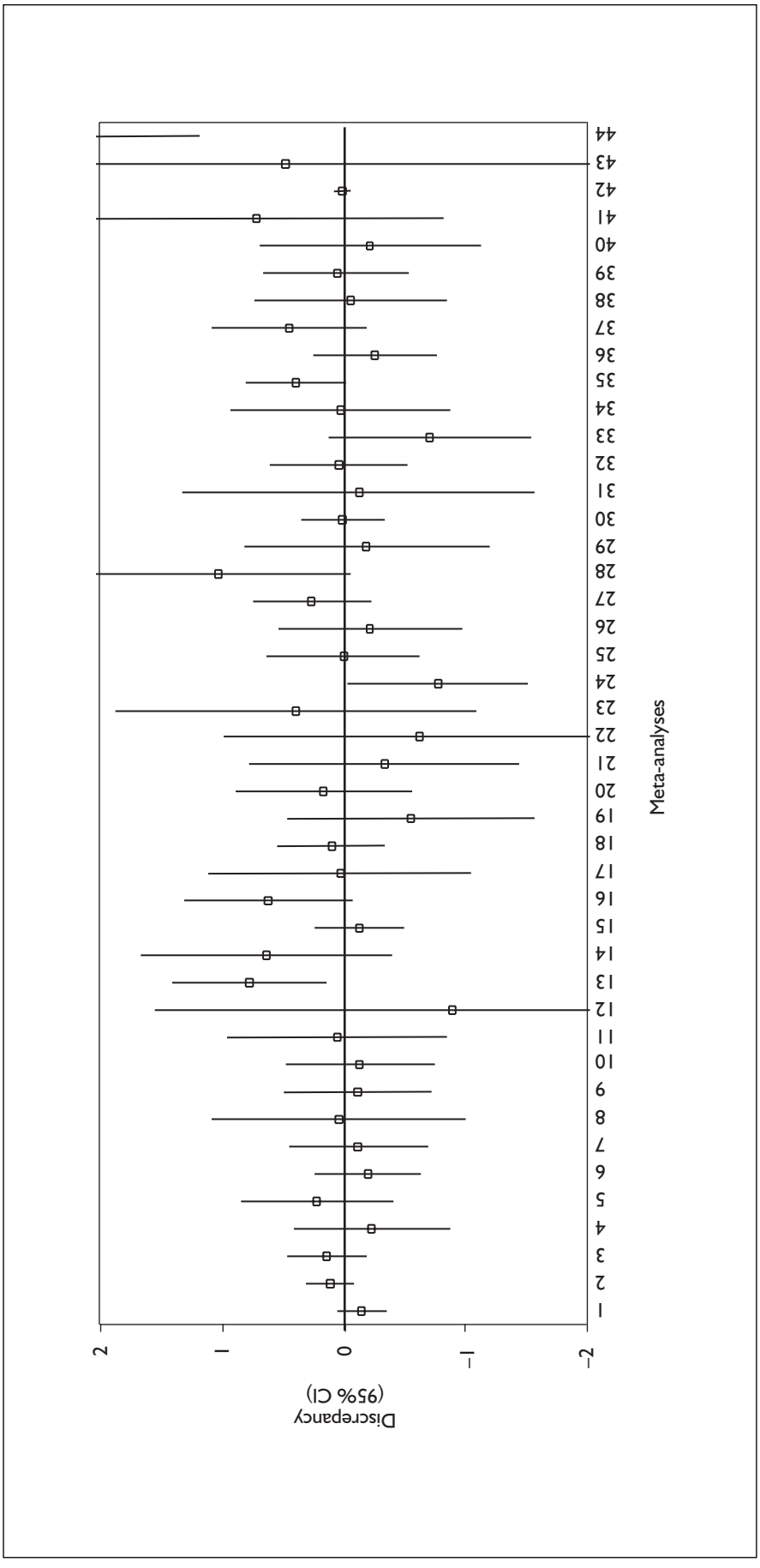

FIGURE 4 *Discrepancy between the direct and the adjusted indirect comparison: empirical evidence from 44 published meta-analyses. Numbering of meta-analyses (1–44) follows the order as in Appendix 8. The discrepancy between the direct and the indirect estimate was defined as difference in estimated log relative risk (meta-analyses 1–39), difference in estimated standardised mean difference (meta-analysis 40) or difference in estimated mean difference (meta-analyses 41–44).*

compared, follow-up period and methods for outcome measurement.

Direct versus naive indirect estimate

Evidence from the naive indirect comparisons can be considered to be equivalent to results from observational or non-randomised controlled research. To explain the discrepancies between the direct and the naive indirect estimates, the most important factor that should be considered is lack of comparability between patients in different studies. The observed difference between treatment groups in the naive indirect comparison may be due not to different interventions of interest, but to different prognostic characteristics among patients in the different studies (confounding).

Random error may be a cause of discrepancy between the direct and the naive indirect estimate. When $|z| = 1.96$ is used as the cut-point for statistical significance, the chance of a type I error is 5%, that is, observing significant discrepancy even if the null hypothesis is true. Because the naive indirect estimates tend to be overprecise (see Chapter 5), the statistical discrepancies between the direct and the naive indirect method will be great, compared with those between the direct and the adjusted indirect method.

Eleven of the 43 comparisons show statistically significant discrepancy between the direct and the naive indirect estimate. In seven of these 11 examples with significant discrepancy, the point estimate by the naive indirect comparison is in the opposite direction to that by the direct comparison. Because of the high frequency of statistically significant discrepancy and unpredictable direction of such discrepancy, the naive indirect method should be avoided in the analysis of data from randomised trials.

Direct versus adjusted indirect estimate

Discrepancies between the direct and the adjusted indirect estimate may also be due to random errors. However, because of the wider confidence interval provided by the adjusted indirect comparison, the discrepancies between the direct and the adjusted indirect estimate are less likely to be statistically significant than those between the direct and the naive indirect estimate. Indeed, the statistically significant results by using the direct comparison often become statistically non-significant in the adjusted indirect comparison (*Table 11*).

More importantly, perhaps, prognostic characteristics of participants in different studies have been taken into account partially by the adjusted indirect method. The adjusted indirect method has partially preserved the rigour of randomisation by considering direct comparisons of interventions of interest with the same control intervention. An underlying assumption in this adjusted method is that the relative efficacy of an intervention is consistent in participants in different studies. It is important to examine the generalisability of results of trials in the adjusted indirect comparison.

Of the 44 comparisons, three showed significant discrepancy ($p < 0.05$) between the direct and adjusted indirect estimates.[27,40,107] These three cases will be further discussed in Chapter 7.

Combination of the direct and the adjusted indirect estimates

It is often the case that direct evidence is available but not sufficient. In such cases, the adjusted indirect comparison may provide supplementary information in evaluating relative efficacy of competing interventions.[53] Sixteen of the 44 direct comparisons included are based on one randomised trial, while the adjusted indirect comparisons were based on a median of 19 trials (range 2–86). Such a large amount of data available for adjusted indirect comparisons could be used to strengthen conclusions based on the direct comparisons, especially when there are concerns about the methodological quality of the single direct comparison trial.

Results of the direct and the adjusted indirect comparison could be quantitatively combined to increase statistical power or precision, when there is no important discrepancy between the two estimates. The statistically non-significant relative effect estimated by the direct comparison may become statistically significant by combining the direct and the adjusted indirect estimate, for example, in two of the 44 meta-analyses (*Figure 5*).

It is also possible that the significant relative effect estimated by the direct comparison would become non-significant after being combined with the adjusted indirect estimate. H_2RA versus sucralfate for non-ulcer dyspepsia from a Cochrane systematic review provides an example.[108] In this example, the direct comparison based on one randomised trial found that H_2RA was less effective than sucralfate [relative risk (RR) 2.74, 95% CI 1.25 to 6.02], while the adjusted indirect comparison indicates that H_2RA was as effective as sucralfate (RR 0.99, 95% CI 0.47 to 2.08). The discrepancy between the direct and the adjusted

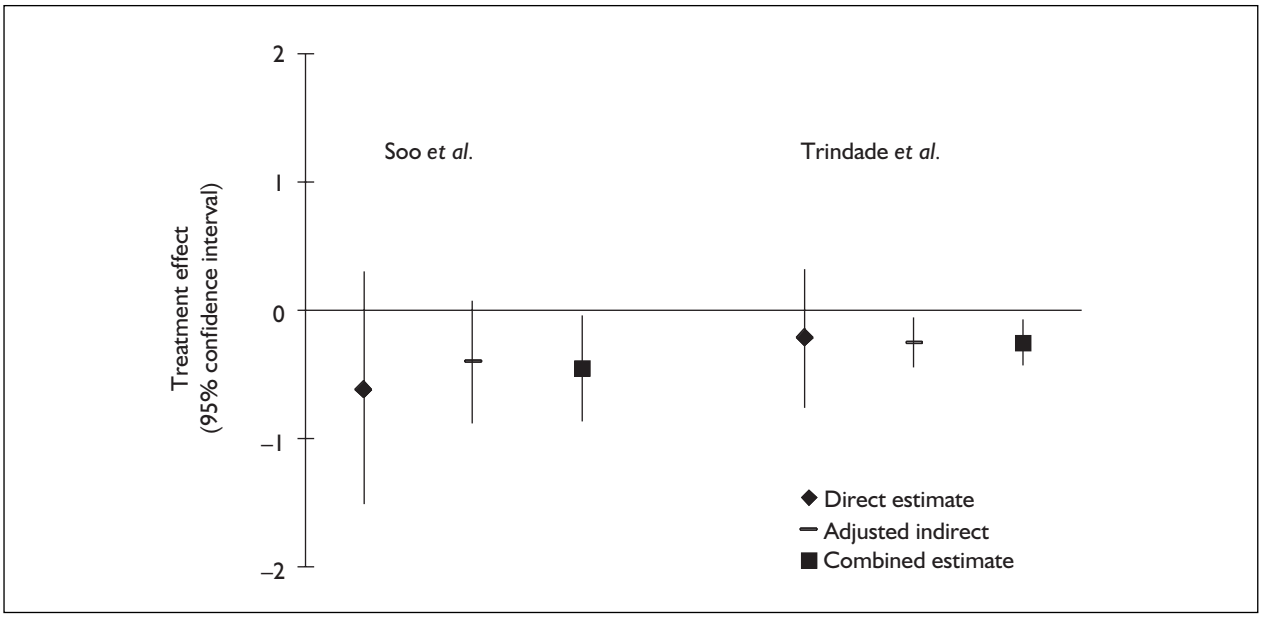

FIGURE 5 *Combining the direct and the adjusted indirect estimates in two meta-analyses*

indirect estimate is statistically significant ($z = 1.84$). The combination of the direct and adjusted indirect estimate provided a statistically non-significant relative risk of 1.56 (95% CI 0.93 to 2.75). In cases like this, one may question whether it is appropriate to combine the two estimates. If it is inappropriate, then one needs to investigate sources of significant discrepancy and determine which estimate is more believable.

Summary

Forty-four analyses from 28 published meta-analyses were available to compare competing interventions both directly and indirectly. There were considerable statistical discrepancies between the direct and the indirect estimate, but the direction of such discrepancy was unpredictable. The relative efficacy may be overestimated or underestimated by the indirect comparison compared with results of the direct comparison.

The naive indirect comparison should be avoided owing to its unpredictable nature and very high frequency of statistically significant discrepancies between the direct and the naive indirect estimate (11/43). In contrast, the adjusted indirect method has two advantages: partially taking into consideration patient prognostic characteristics and wider confidence intervals. Empirical evidence presented in this chapter has confirmed these theoretical advantages. There is no statistically significant discrepancy between the direct and the adjusted indirect estimate in most cases (41/44). When direct evidence is available but not sufficient, the direct and the adjusted indirect estimate could be combined to obtain a more precise estimate.

Chapter 7

Detailed case studies

This chapter first examines five comparisons from the sample of meta-analyses included in Chapter 6.[27,40,107–109] The reason for selecting the five cases is that they showed a statistically significant discrepancy ($z > 1.64$) between the direct and the adjusted indirect estimate. It may be interesting to note that four of these five examples were about the treatment with H_2RA or PPI.[40,107–109]

Following this, further empirical evidence is provided about indirect comparisons, using a systematic review of antimicrobial prophylaxis in colorectal surgery.[3]

Five cases with significant discrepancy

Case 1: Chiba and colleagues (1997)[40]

This meta-analysis evaluated speed of healing and symptom relief in grade II–IV GORD. The review included only randomised trials but did not directly compare different interventions, although it has presented sufficient data for both direct and indirect comparison of H_2RA and PPI.

By pooling data from 13 trials that compared H_2RA and PPI directly, the healing proportion in patients receiving H_2RA was lower than that in patients receiving PPI (RR 0.56, 95% CI 0.48 to 0.66). The adjusted indirect comparison between H_2RA and PPI was carried out using 11 trials that compared H_2RA versus placebo and two trials that compared PPI versus placebo (*Figure 6*). The adjusted indirect estimate (RR 0.26, 95% CI 0.14 to 0.48) was statistically significantly different from the direct estimate ($z = 2.35$). However, there may be no important clinical implication associated with this statistically significant discrepancy. The discrepancy may be quantitative rather than qualitative since both the direct and adjusted indirect estimates are in the same direction. In this example, the naive indirect estimate is also statistically significantly different from the direct estimate ($z = –2.95$) (*Figure 6a*).

Figure 6(b) presents results of each intervention group from trials used in the direct and indirect comparisons of H_2RA and PPI. Patients receiving placebo had worse outcome in two PPI trials (12.0%) than those receiving placebo in 11 H_2RA trials (35.3%). One explanation for this observation may be that the patients in the two PPI trials were more severe than those in the H_2RA trials. However, further investigation is impossible because detailed data on patient characteristics were not available.

Case 2: Rostom and colleagues (2000)[107]

This systematic review assessed the effectiveness of interventions for the prevention of NSAID-induced upper gastrointestinal toxicity. The interventions of interest include misoprostol, H_2RA and PPI. Although the authors of this review did not make any indirect comparisons, the data from the individual studies are sufficient to compare PPI and H_2RA both directly and indirectly. In a trial that directly compared PPI and H_2RA in reducing total endoscopic ulcers, significant difference in favour of PPI was observed (RR 0.28, 95% CI 0.15 to 0.51). The adjusted indirect comparison was conducted using three trials that compared PPI versus placebo and six trials that compared H_2RA versus placebo. Significant discrepancy was observed between the direct estimate and the adjusted indirect estimate, and between the direct and the naive indirect estimate (*Figure 7*).

In this review the authors concluded that low-dose H_2RA was less effective than high-dose H_2RA. Because low dose H_2RA was used in the trial of direct comparison, the indirect comparisons were also conducted after separating H_2RA trials into subsets of high dose and low dose (according to authors' original definition). Results of the sensitivity analysis indicate that the discrepancy between the direct and the adjusted indirect estimate was reduced but still statistically significant when low-dose H_2RA was compared with PPI. The naive indirect estimate in this example is also significantly different from the direct estimate. According to the naive indirect estimate, there is no significant difference in total endoscopic ulcers between the PPI and H_2RA treatment.

Figure 7(b) presents the results of each intervention group from the trials included. It can be seen that

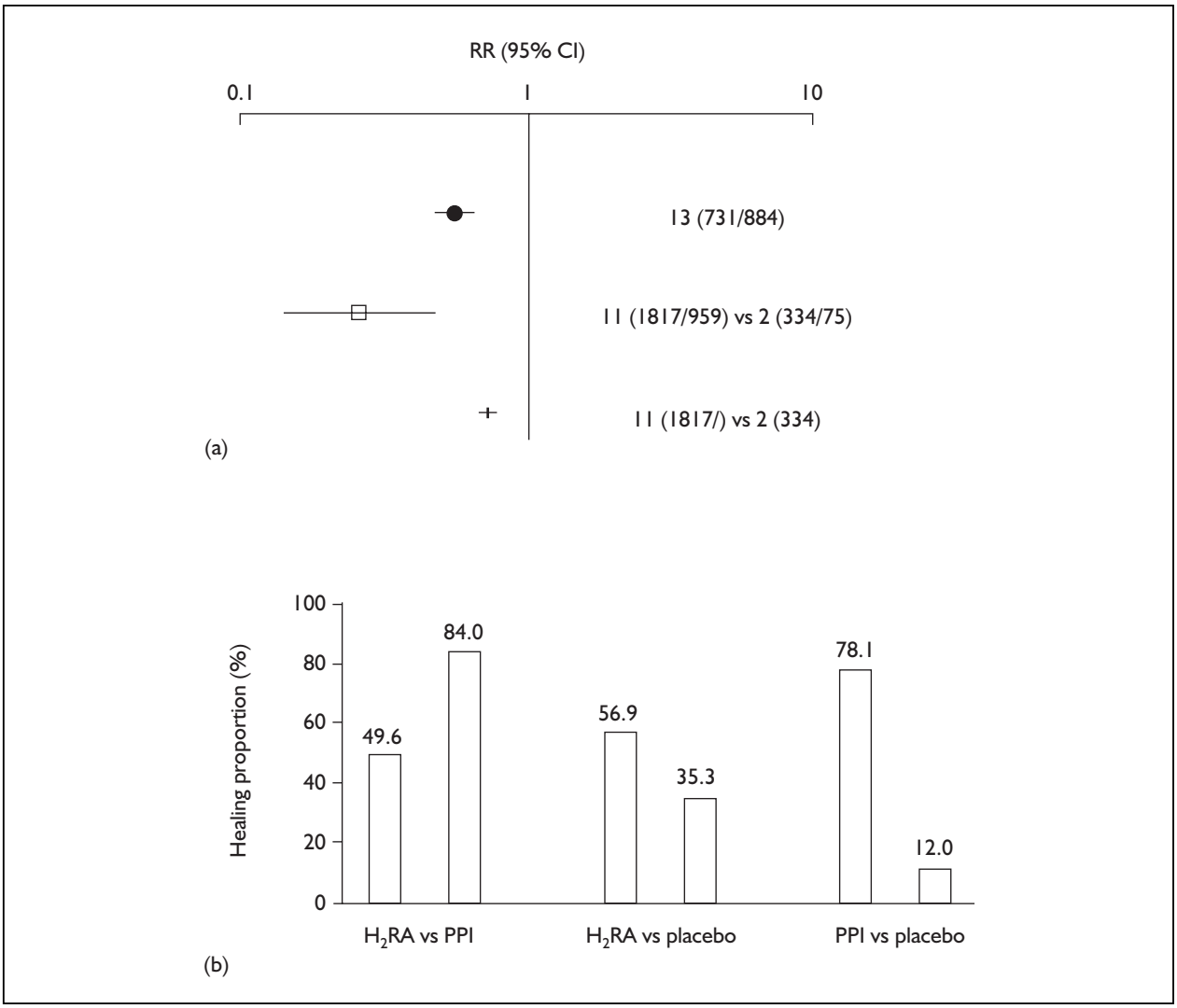

FIGURE 6 *Chiba* et al.:[40] *H_2RA versus PPI for GORD (healing proportion). (a) Solid circle, direct estimate; square, adjusted indirect; cross, naive indirect. The numbers shown in the figure are the number of trials (patients). (b) Results of each intervention group from trials used in the direct and indirect comparison of H_2RA and PPI.*

patients receiving PPI had better outcome while patients receiving H_2RA had worse outcome in the trial that directly compared PPI and H_2RA, compared with patients receiving the same intervention in placebo-controlled trials. *Table 12* presents some characteristics of studies involved in the direct and indirect comparisons. It seems that patients in PPI trials and in H_2RA trials were similar. The results of individual trials and meta-analyses are shown in *Figure 8*. There was no significant heterogeneity across trials for any set of trials. Thus, the causes of the statistically significant discrepancy were unknown.

The observed statistical discrepancy between the direct and adjusted indirect estimate may not be clinically important in this case. The results of the direct and adjusted indirect method both suggested that PPI was superior to H_2RA. Since the direct estimate was based on only one trial that was not double blinded, the relative efficacy of PPI versus H_2RA in preventing NSAID-induced endoscopic ulcers may not be as great as had been suggested by the direct estimate.

Case 3: Soo and colleagues (2000)[108]

This Cochrane systematic review evaluated pharmacological interventions for non-ulcer dyspepsia. The authors attempted indirect comparisons of different drugs and used multiple regression in these (adjusted) indirect comparisons. Only one comparison had sufficient data for both a direct and an adjusted indirect estimate, and this is discussed below.

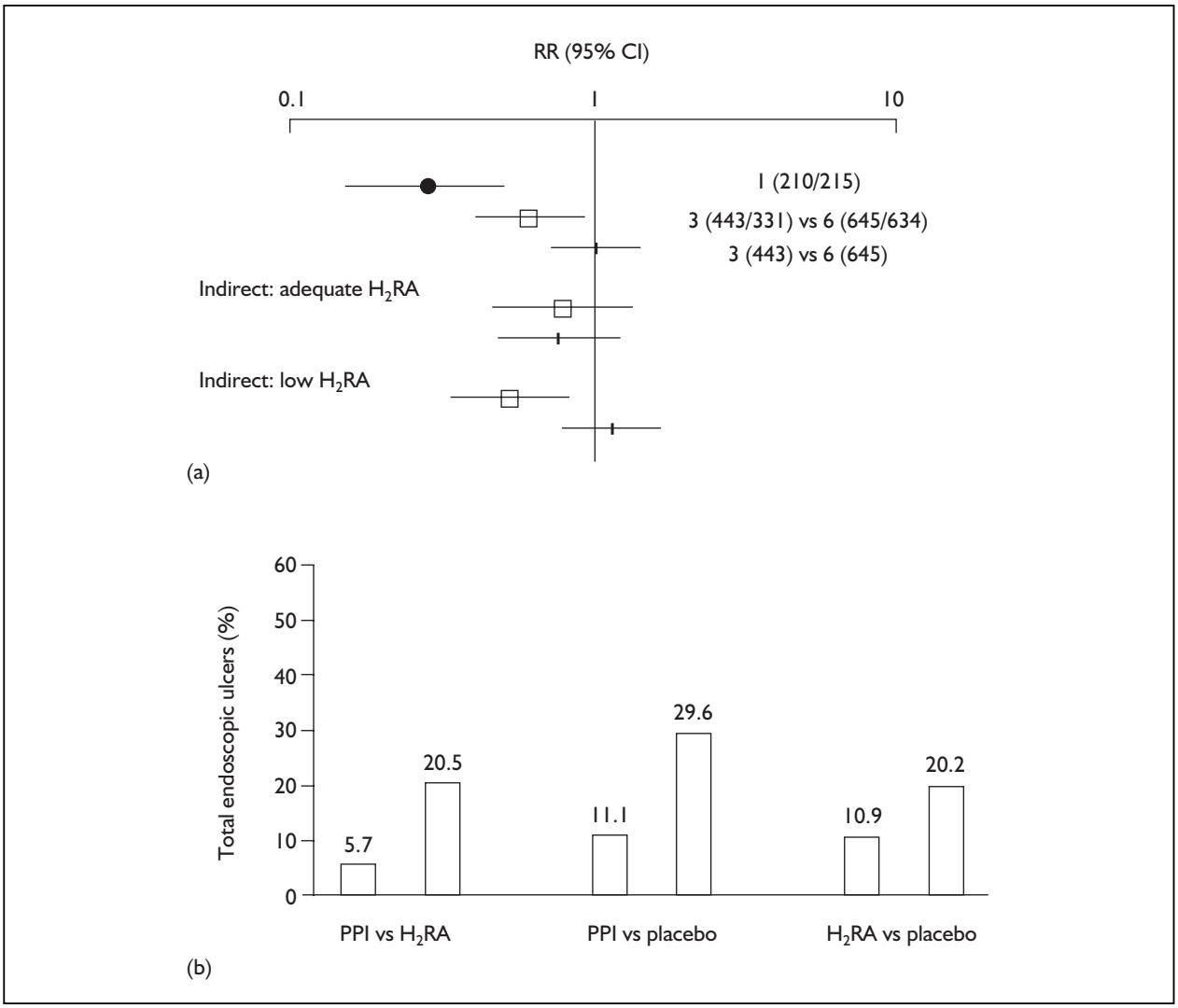

FIGURE 7 *Rostom* et al.:[107] *PPI versus H_2RA for preventing chronic NSAID-induced upper gastrointestinal toxicity (endoscopic ulcers). (a) Solid circle, direct estimate; square, adjusted indirect; cross, naive indirect. The numbers shown in the figure are the number of trials (patients). (b) Results of each intervention group from trials used in the direct and indirect comparison of H_2RA and PPI.*

One trial compared H_2RA and sucralfate directly and observed a significant difference in global symptom assessment between the two interventions (RR 2.74, 95% CI 1.25 to 6.02). The adjusted indirect comparison involved eight placebo-controlled H_2RA trials and two placebo-controlled sucralfate trials (*Figure 9*). The adjusted indirect estimate indicates no difference between H_2RA and sucralfate (RR 0.99, 95% CI 0.47 to 2.08). The discrepancy between the direct estimate and the adjusted indirect estimate is statistically significant ($z = 1.85$) (*Figure 10*).

Figure 10(b) indicates that the result of sucralfate arm in the direct trial is extremely good compared with trials used in the adjusted indirect comparison. This cannot be explained easily by factors such as different underlying risk because the result of H_2RA arm in the direct trial was similar to that in the trials involved in the adjusted indirect comparison. The scrutiny of characteristics of studies in *Table 13* provided no obvious explanation for the observed discrepancy between the direct and indirect estimates and heterogeneity among trials. In this case, the discrepancy between the direct and adjusted indirect comparison was clinically important. The result of the direct comparison by the small and open trial may not be more believable than the result of the adjusted indirect comparison.

Case 4: van Pinxteren and colleagues (2000)[109]

This Cochrane systematic review evaluated short-term treatment for GORD-like symptoms. For the comparison of empirical treatment of heartburn

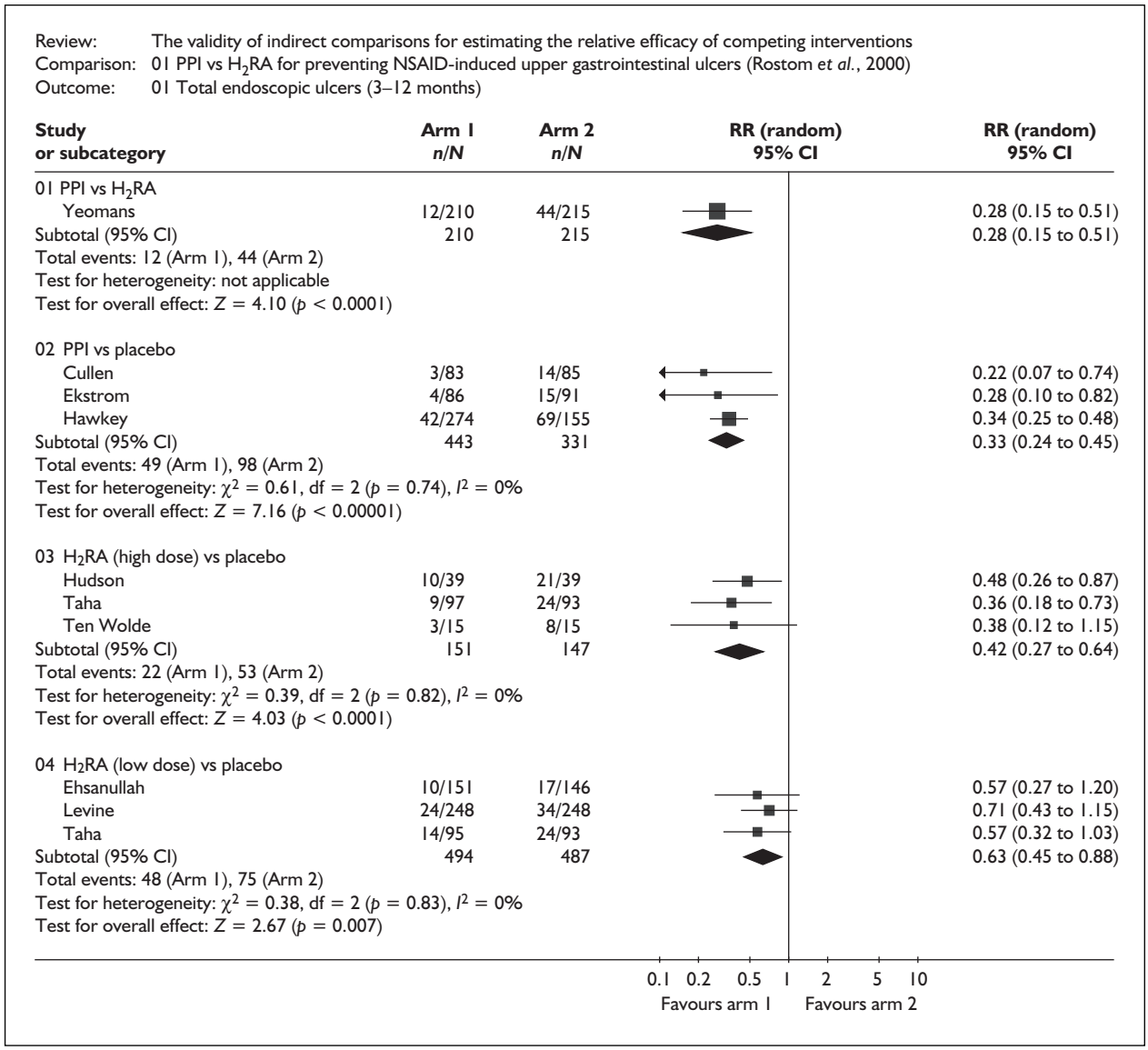

FIGURE 8 *Results of individual trials involved in the direct and indirect comparisons in the meta-analysis by Rostom* et al.[107]

remission with PPI or H_2RA, the direct estimate (RR 0.67, 95% CI 0.57 to 0.80) was statistically significantly different from the adjusted indirect estimate (RR 0.45, 95% CI 0.31 to 0.66). This statistical discrepancy may have no actual clinical importance in this review, since both estimates indicated a significant benefit of PPI versus H_2RA in the empirical treatment for heartburn remission (*Figure 11*).

Case 5: Zhang and Li-Wan-Po (1996)[27]

This meta-analysis evaluated the analgesic efficacy of paracetamol plus codeine in surgical pain. (The author, WY Zhang, supplied data for individual studies that were not available in the published article.) The direct comparison indicated that paracetamol plus codeine was more efficacious that paracetamol alone (mean difference in the sum of pain intensity difference 6.97, 95% CI 3.56 to 10.37). The adjusted indirect comparison did not show significant difference between paracetamol plus codeine and paracetamol alone (mean difference –1.16, 95% CI –6.95 to 4.64) (*Figure 12*).

Results of each intervention group from trials involved in the direct and the indirect comparisons are presented in *Figure 12(b)*. The improvement in pain intensity was much less in patients receiving placebo in placebo-controlled trials of paracetamol plus codeine than in placebo-controlled trials of paracetamol alone. It suggests that patients in the placebo-controlled trials of paracetamol plus codeine may be different to

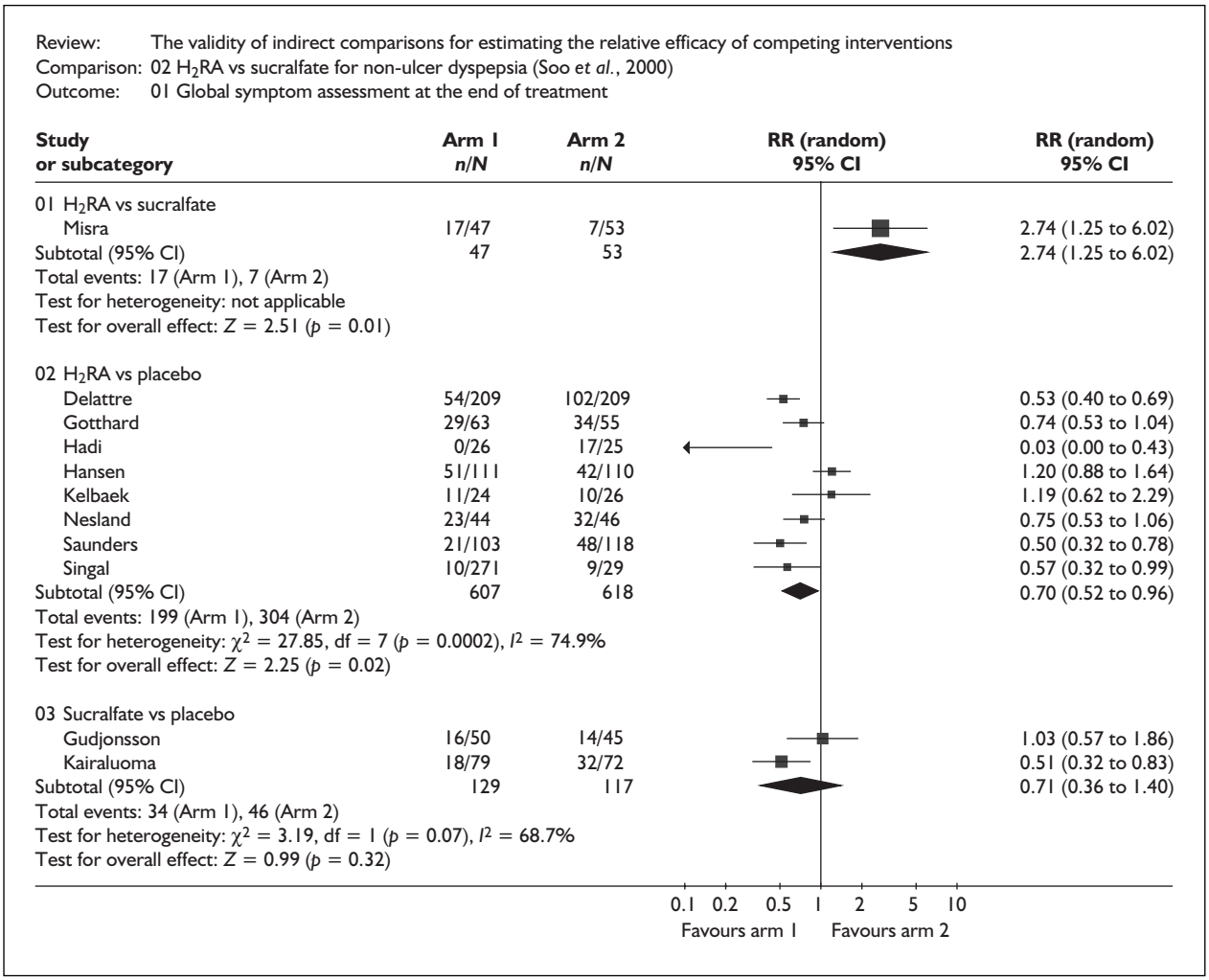

FIGURE 9 *Results of individual trials involved in the direct and indirect comparisons in the meta-analysis by Soo* et al.[108]

those in the placebo-controlled trials of paracetamol alone. The possible difference in baseline risk is taken into consideration in the adjusted indirect comparison, which indicates that there is no difference between paracetamol plus codeine and paracetamol alone. When the different baseline risk has not been considered, the naïve indirect comparison yields a result that is opposite to the direct estimate (*Figure 12a*).

In this case, the significant discrepancy may be due to different types of surgical pain and/or different doses of paracetamol/codeine used in different trials. When the analysis was restricted to trials of dental surgery, the discrepancy between the direct and the adjusted indirect estimate was still significant (*Table 14*). Further scrutiny of the included trials found that the majority of the trials ($n = 10$) in the direct comparison used a dose of 600–650 mg for paracetamol and 60 mg for codeine, whereas many placebo-controlled trials ($n = 29$) used a dose of 300 mg for paracetamol and 30 mg for codeine. When the analysis included only trials that used paracetamol 600–650 mg and codeine 60 mg, the adjusted indirect estimate (5.72, 95% CI –5.37 to 16.81) was no longer significantly different from the direct estimate (7.28, 95% CI 3.69 to 10.87). Thus, the significant discrepancy between the direct and the indirect estimate based on all trials could be explained by the fact that many placebo-controlled trials used low doses of paracetamol (300 mg) and codeine (30 mg). This example suggests that the similarity of trials involved in the adjusted indirect comparison should be carefully assessed.

Relative efficacy of antimicrobial prophylaxis in colorectal surgery

(Note: this section is based on an article published in *Controlled Clinical Trials*.[87])

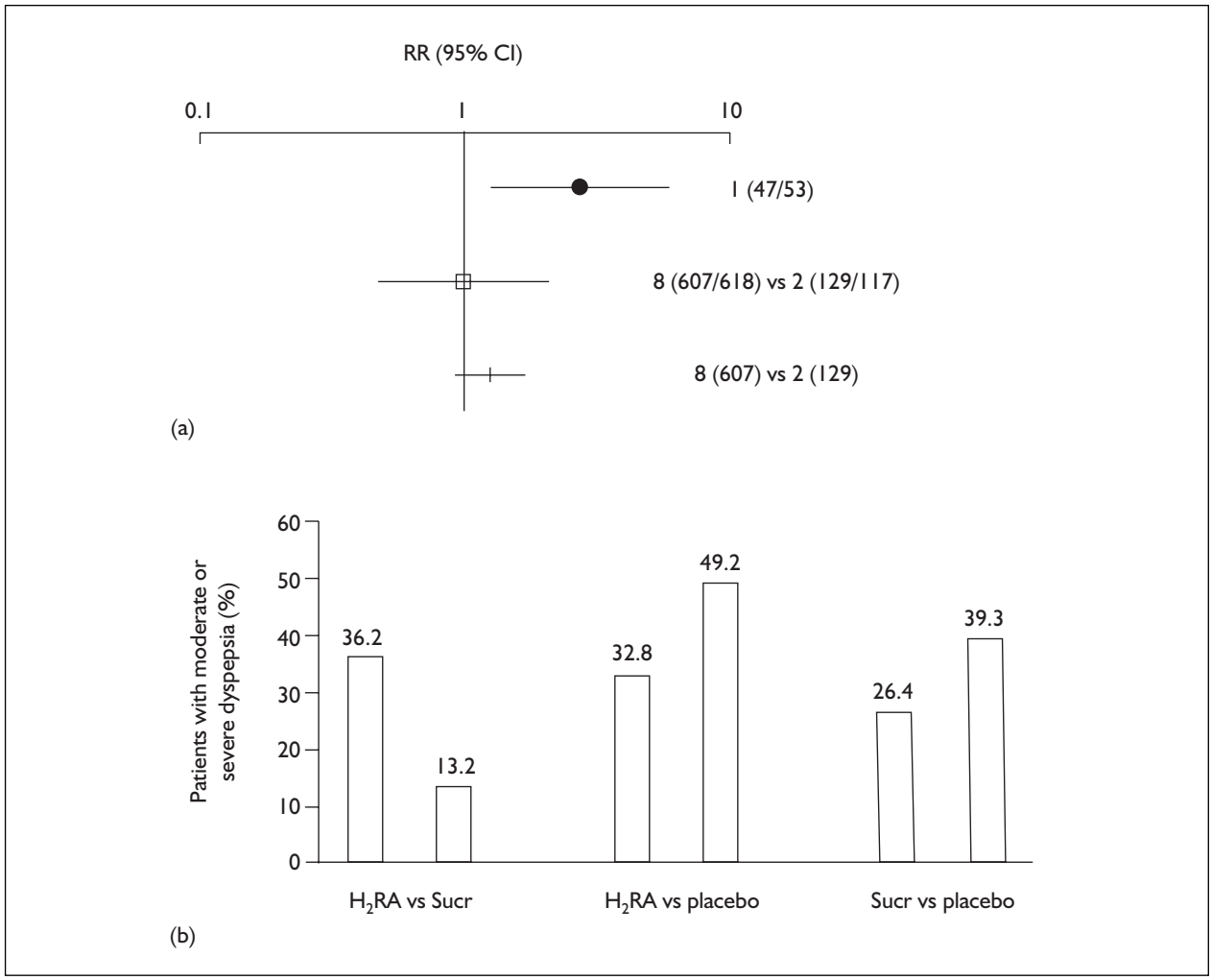

FIGURE 10 *Soo* et al.*:[108] H_2RA versus sucralfate (Sucr) for non-ulcer dyspepsia (global symptom assessment). (a) Solid circle, direct estimate; square, adjusted indirect; cross, naive indirect. The numbers shown in the figure are the number of trials (patients). (b) Results of each intervention group from trials used in the direct and indirect comparison of H_2RA and sucralfate.*

In a systematic review of antibiotic prophylaxis for preventing surgical wound infection after colorectal surgery,[3] 147 randomised trials were identified in which more than 70 different antibiotics or combinations of antibiotics were assessed. Only a limited number of antibiotics was directly compared within the trials. If all 70 options had been directly compared, over 2400 trials would have been required, without considering different dosages, routes and timing of administration of the same drug.

Using this systematic review as an example, this section aims to explore the potential usefulness and limitations of indirect comparison in evaluating the relative efficacy of competing interventions.

Method

From the systematic review of antimicrobial prophylaxis in colorectal surgery 11 sets of trials were identified in which different antibiotics could be compared both directly and indirectly.[3] Each set of trials contains at least three trials that tested three different antibiotics (or combinations of antibiotics). For example, suppose a set of trials includes a trial that compared antibiotic A with B, a trial that compared A with C, and a trial that compared B with C. Then the trials in this set could be used for three different comparisons: A versus B, A versus C and B versus C.

For each comparison the relative efficacy of antimicrobial prophylaxis (i.e. odds ratio for surgical wound infection) was estimated using three different methods: direct comparison, adjusted indirect comparison and naive indirect comparison. The method suggested by Bucher and colleagues was used to carry out the adjusted indirect comparison.[14] Results from more than one trial were weighted by the inverse of corresponding variances and then quantitatively

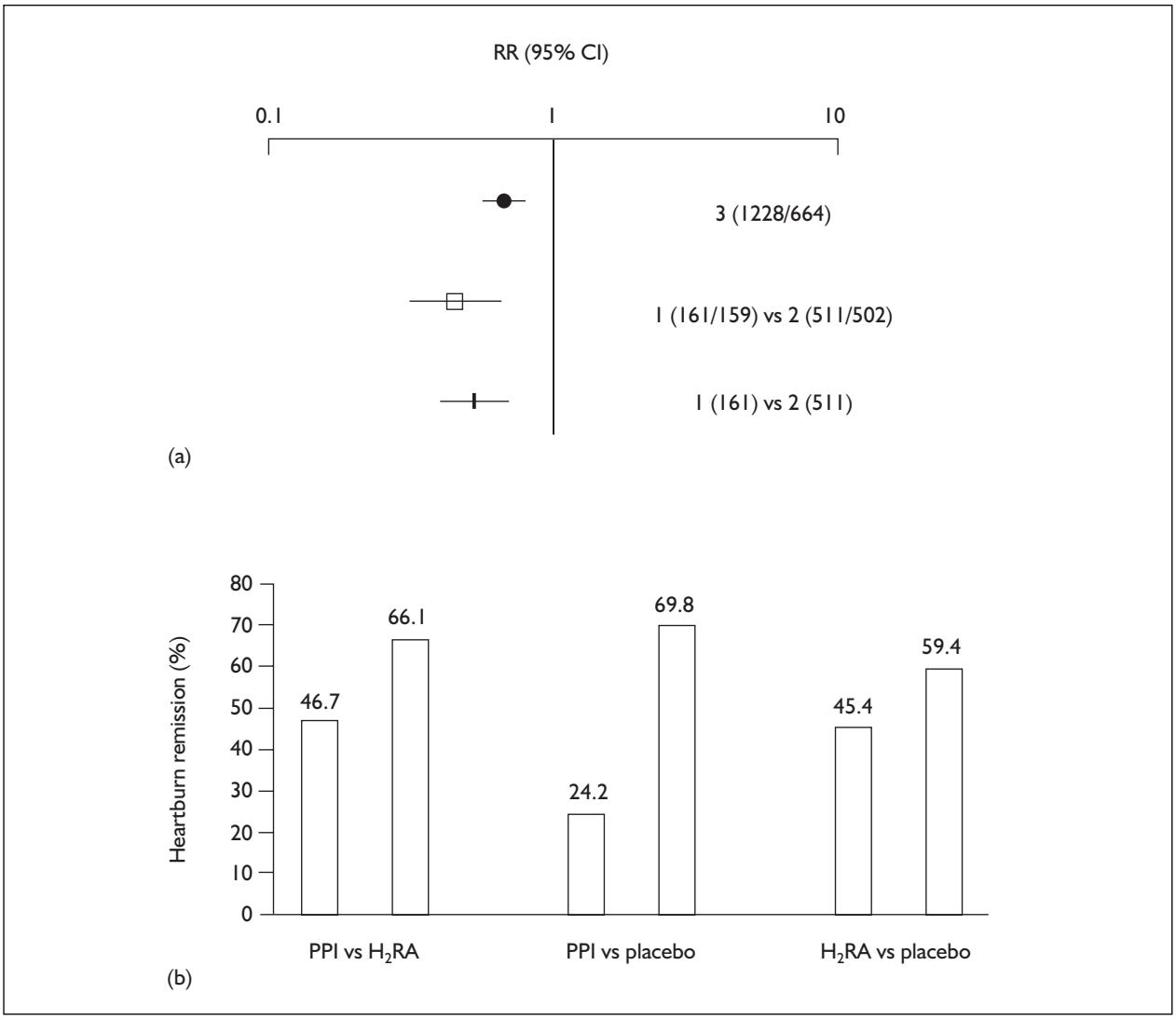

FIGURE 11 *van Pinxteren* et al.:[109] *PPI versus H_2RA for reflux disease-like symptoms (heartburn remission). (a) Solid circle, direct estimate; square, adjusted indirect; cross, naive indirect. The numbers shown in the figure are the number of trials (patients). (b) Results of each intervention group from trials used in the direct and indirect comparison of H_2RA and PPI.*

pooled to obtain an overall estimate. The naive indirect method compared the results of single arms included in the trials that had been used in the adjusted indirect comparison. In the naive indirect comparison, when relevant the results of more than one trial were pooled by adding up the number of surgical wound infections and the number of patients.

As an example, *Table 15* presents data from a set of three trials that could be used to conduct three different comparisons: cefuroxime plus metronidazole (Cefur-M) versus co-amoxiclav (Co-A),[110] cefuroxime plus metronidazole (Cefur-M) versus cefotaxime plus metronidazole (Cefot-M)[111] and co-amoxiclav versus cefotaxime plus metronidazole.[112] *Table 16* shows the results of different analyses for each of the three comparisons. For instance, Cefur-M and Co-A were directly compared in the trial conducted by Palmer and colleagues.[110] The indirect comparison of Cefur-M and Co-A was based on the other two trials that included the common intervention Cefot-M.[111,112] Likewise, the direct comparison of Cefur-M and Cefot-M was based on the results obtained by Rowe-Jones and colleagues,[111] and the indirect comparison was based on the two trials with a common intervention, Co-A.[110,112]

The results of the two indirect methods were compared with the results from the gold standard of direct comparison. In this example, the discrepancy was defined as the absolute value of difference in log odds ratio between the direct method and the indirect method. To take into

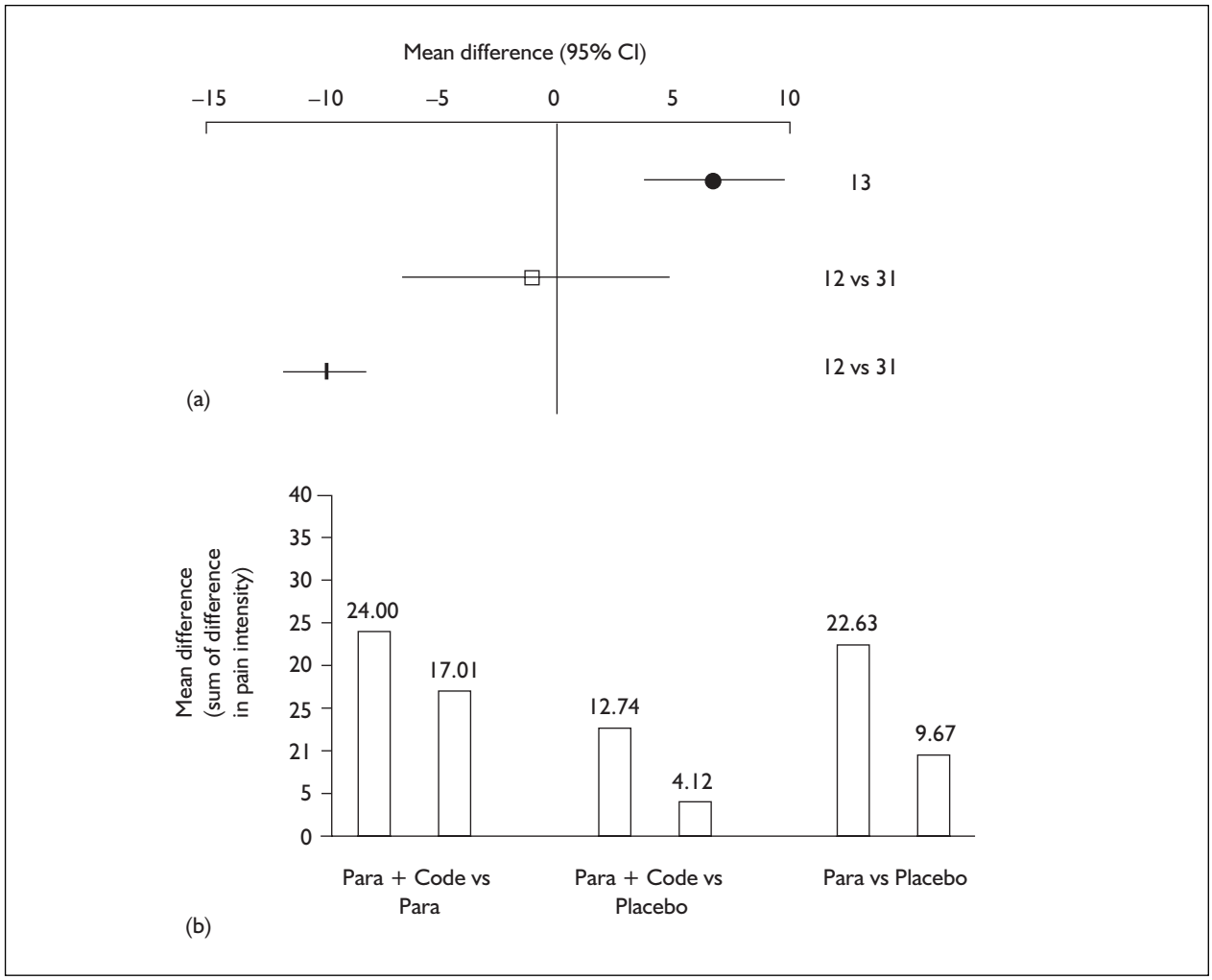

FIGURE 12 *Zhang and Li Wan Po:[27] paracetamol (Para) plus codeine (Code) versus paracetamol alone in surgical pain (analgesic efficacy). (a) Solid circle, direct estimate; square, adjusted indirect; cross, naïve indirect. The numbers shown in the figure are the number of trials (patients). (b) Results of each intervention group from used in the direct and indirect comparison of paracetamol plus codeine and paracetamol alone.*

account the precision of estimated discrepancy, the corresponding z statistic was calculated by dividing the difference in log odds ratio by the square root of the sum of the variances. The cut-off point for statistical significance was arbitrarily considered to be $z = 1.64$ (or two-sided p-value = 0.10).

The difference in log odds ratio can be converted into the ratio of odds ratios (i.e. the antilogarithm of difference in log odds ratio). For example, when the difference in log odds ratio is 1.0, the corresponding ratio of odds ratios is 2.72. It can be mathematically shown that the absolute value of discrepancy between the adjusted indirect method and the direct method will be the same for three different comparisons using the same set of trials. For example, it was 1.636 for all three comparisons in *Table 16* (this example corresponds to trial set 5 in *Figure 13*).

Results

The results presented in *Figure 13* indicate that considerable discrepancies exist between the direct and indirect comparisons. In each set of trials, the discrepancies between the direct and the adjusted indirect method were the same in three different comparisons. In contrast, the discrepancies between the direct and the naive indirect comparisons were unpredictable and varied greatly across the different comparisons.

The discrepancies between the direct and the naive indirect comparisons may be either smaller or greater than those between the direct and the adjusted indirect comparisons (*Figure 13*), depending on which interventions have been compared using a given set of trials. In nine of the 11 sets of trials, there is a naive indirect comparison with a greater discrepancy than the adjusted indirect comparison. The significant

TABLE 12 *Characteristics of studies involved in the direct and the indirect comparisons of PPI and H_2RA for preventing NSAID-induced upper gastrointestinal toxicity (total endoscopic ulcers)*[107]

Study	Intervention (duration)	Participants	Quality[a]
Yeomans, 1998	Omeprazole 20 mg per day Ranitidine 2× 150 mg (6 months)	Patients with RA and OA who needed maintenance treatment after successful treatment of ulcers or erosions in either the stomach or duodenum Mean age 56 years *Heliconacter pylori*: positive: 50%	Jadad scale: 3 Allocation concealment: D
Cullen, 1998	Omeprazole 20 mg per day Placebo (6 months)	Patients taking NSAIDs regularly, chronically and above defined minimum doses	Jadad scale: 2 Allocation concealment: D
Ekstrom, 1996	Omeprazole 20 mg per day Placebo (3 months)	Patients with dyspepsia or history of peptic ulcer disease Mean age 58 years Previous peptic ulcer: 27% *Helicobacter pylori* positive: 53%	Jadad scale: 3 Allocation concealment: D
Hawkey, 1998	Omeprazole 20 mg Misoprostol 400 μg Placebo (6 months)	Patient with RA and OA. Maintenance therapy for patients in whom treatment was successful Outpatients Mean age 58 years Length of NSAIDs: > 6 months Previous peptic ulcers: 29% *Helicobacter pylori* positive: 42%	Jadad scale: 3 Allocation concealment: D
Hudson, 1997	Famotidine 40 mg b.d. (high dose) Placebo (6 months)	Patients with RA or OA. NSAID users with healed ulcers *Helicobacter pylori* positive: 18/39	Jadad scale: 3 Allocation concealment: D
Taha, 1996	Famotidine 20 mg b.d. (low dose) Famotidine 40 mg b.d. (high dose) Placebo (6 months)	Patients without peptic ulcers who were receiving long-term NSAID therapy for RA (82%) or OA	Jadad scale: 4 Allocation concealment: D
Ten Wolde, 1996	Ranitidine 300 b.d. (high dose) Placebo (The study was stopped after a blinded interim analysis)	Patients with RA and a history of peptic ulcer disease	Jadad scale: 3 Allocation concealment: D
Ehsanullah, 1988	Ranitidine 150 mg b.d. (low dose) Placebo (2 months)	Patients with RA or OA aged >18 years without lesions in the stomach and duodenum at baseline endoscopy (after 1 week without taking NSAIDs). Those taking other antirheumatic agents, concomitant ulcerogenic drugs or treatment for peptic ulcers within the previous 30 days were excluded Mean age 57 years	Jadad scale: 5 Allocation concealment: D
Levine, 1993	Nizatidine 150 mg b.d. Placebo (3 months)	Patients with OA who were taking NSAIDs. Endoscopy to rule out the presence of an acute ulcer	Jadad scale: 3 Allocation concealment: D

[a] As coded in the original review.[107]
OA, osteoarthritis; RA, rheumatoid arthritis.

TABLE 13 *Characteristics of studies involved in the direct and the indirect comparisons of H_2RA and sucralfate for non-ulcer dyspepsia*[108]

Study	Interventions	Participants	Quality[a]
Misra, 1992 (India)	Ranitidine 150 mg Sucralfate 1 g q.d. (4 weeks)	Patients with 1 month of abdominal symptoms referable to the upper GI tract. Patients with GORD or IBS were excluded. 87% completed trial	Randomised, open, controlled clinical trial Allocation concealment: B
Delattre, 1985 (USA)	Cimetidine 200 mg q.d. Placebo (4 weeks)	Patients with non-ulcer dyspepsia	RCT, double-blind, placebo-controlled, multicentre trial Allocation concealment: B
Gotthard, 1988 (Sweden)	Cimetidine 400 mg b.d. Antacid 10 ml q.d. Placebo (6 weeks)	3/12 of dyspepsia of unknown origin. Acid output studies performed. 16% duodenitis	RCT, double-blind, placebo-controlled trial Allocation concealment: B
Hadi, 1989 (Indonesia)	Ranitidine 300 mg daily Placebo (4 weeks)	Duration of dyspepsia unclear. Gastritis on all OGD. Dropout rate for placebo was 23% and for ranitidine was 4%	RCT, double-blind, placebo-controlled trial Allocation concealment: B
Hansen, 1998 (Denmark)	Cisapride 10 mg t.d. Nizatidine 300 mg nocte Placebo (2 weeks)	Primary care recruitment. Mean duration of dyspeptic symptom was 88 months. Four subgroups: ulcer-like (13%), reflux-like (23%), dysmotility-like (46%) and unclassified (18%). Included superficial erosions on OGD. 85% completed trial	RCT, double-blind, placebo-controlled trial Allocation concealment: A
Kelbaek, 1985 (Denmark)	Cimetidine 200 mg t.d. and 400 mg nocte Placebo (3 weeks)	Primary care recruitment. Patients with 1 month of epigastric pain. Acid output studies performed. Had OGD. 14 patients who were symptom free at end of treatment had 3 months of follow-up. 96% completed trial	RCT, double-blind, placebo-controlled trial Allocation concealment: B
Nesland, 1985 (Norway)	Cimetidine 400 mg b.d. Placebo (4 weeks)	Patients with 6 months of predominantly ulcer-like pain with erosive prepyloric changes. 90% completed trial	RCT, double-blind, placebo-controlled trial Allocation concealment: B
Saunders, 1986 (UK)	Ranitidine 150 mg b.d. Placebo (6 weeks)	Primary care recruitment. 88% completed trial. One-year follow-up, but the results included other peptic disease	RCT, double-blind, placebo-controlled multicentre trials Allocation concealment: A
Singal, 1989 (India)	Cimetidine 400 mg b.d. Placebo (4 weeks)	Patients with 1 month of primary symptom of upper abdominal discomfort. IBS excluded	RCT, double-blind, placebo-controlled trial Allocation concealment: B
Gudjonsson, 1993 (Iceland)	Sucralfate 1 g q.d. Placebo (3 weeks)	Private practice recruitment. Patients with 1 month of dyspepsia. Had OGD. 91% completed trial	RCT, double-blind, placebo controlled trial Allocation concealment: B
Kairaluoma, 1987 (Finland)	Sucralfate 1 g t.d. Placebo (4 weeks)	Patients with 3 months of dyspepsia. Had OGD. 6% had duodenitis. 86% completed trial	RCT, double-blind, placebo-controlled trial Allocation concealment: D

[a] As coded in the original review[108]

GI, gastrointestinal; IBS, irritable bowel syndrome; OGD, oesophagogastroduodenoscopy.

TABLE 14 *Direct and adjusted indirect estimates of efficacy of paracetamol plus codeine versus paracetamol alone for surgical pain*[27]

Trials	Mean difference in standardised score for the sum of pain intensity difference (95% CI)		Discrepancy between the direct and adjusted indirect estimates
	Direct comparison	Adjusted indirect comparison	
All trials	6.97 (3.56 to 10.37) $n = 13$	–1.16 (–6.95 to 4.64) $n_1 = 12$, $n_2 = 31$	8.13 ($p = 0.018$)
Dental surgery trials only	7.07 (3.37 to 10.78) $n = 11$	–1.40 (–8.27 to 5.46) $n_1 = 7$, $n_2 = 15$	8.47 ($p = 0.033$)
Trials that used paracetamol 600–650 mg and codeine 60 mg	7.28 (3.69 to 10.87) $n = 10$	5.72 (–5.37 to 16.81) $n_1 = 2$, $n_2 = 12$	1.56 ($p = 0.793$)

n, n_1 and n_2 are the number of trials used in the direct comparison and indirect comparison, respectively.
Trials included dental surgery, episiotomy, postpartum uterine cramp, orthopaedic and other surgery. Doses of paracetamol ranged from 300 to 1000 mg and the dose of codeine was 30 or 60 mg.

TABLE 15 *Example: a set of trials that could be used to compare different antibiotic regimens both directly and indirectly for preventing surgical wound infection in colorectal surgery*

Trial	Cefur-M SWI/*n*	Co-A SWI/*n*	Cefot-M SWI/*n*
Palmer *et al*., 1994[110]	2/79	8/69	–
Rowe-Jones *et al*., 1990[111]	33/454	–	32/453
Kwok *et al*., 1993[112]	–	7/76	8/88

n, number of patients; SWI, surgical wound infection.

TABLE 16 *Example: results of different methods for each comparison between antibiotics, using the trials presented in* Table 15

Comparison	Direct method lnOR (95% CI)	Naive indirect method lnOR (95% CI)	Adjusted indirect method lnOR (95% CI)
Cefur-M vs Co-A	–1.619 (–3.205 to –0.034)	–0.258 (–1.112 to 0.596) $\Delta = 1.361$ $z = 1.481$	0.016 (–1.162 to 1.194) $\Delta = 1.636$ $z = 1.623$
Cefur-M vs Cefot-M	0.031 (–0.474 to 0.535)	–1.348 (–2.929 to 0.233) $\Delta = -1.379$ $z = -1.629$	–1.605 (–3.514 to 0.305) $\Delta = -1.636$ $z = -1.623$
Co-A vs Cefot-M	0.0144 (–1.050 to 1.079)	0.545 (–0.275 to 1.365) $\Delta = 0.531$ $z = 0.775$	1.650 (–0.014 to 3.314) $\Delta = 1.636$ $z = 1.623$

lnOR, log odds ratio; Δ, difference in log odds ratio between the indirect method and the direct method.
The z statistic was calculated by dividing the difference in log odds ratio by the square root of the sum of the variances.

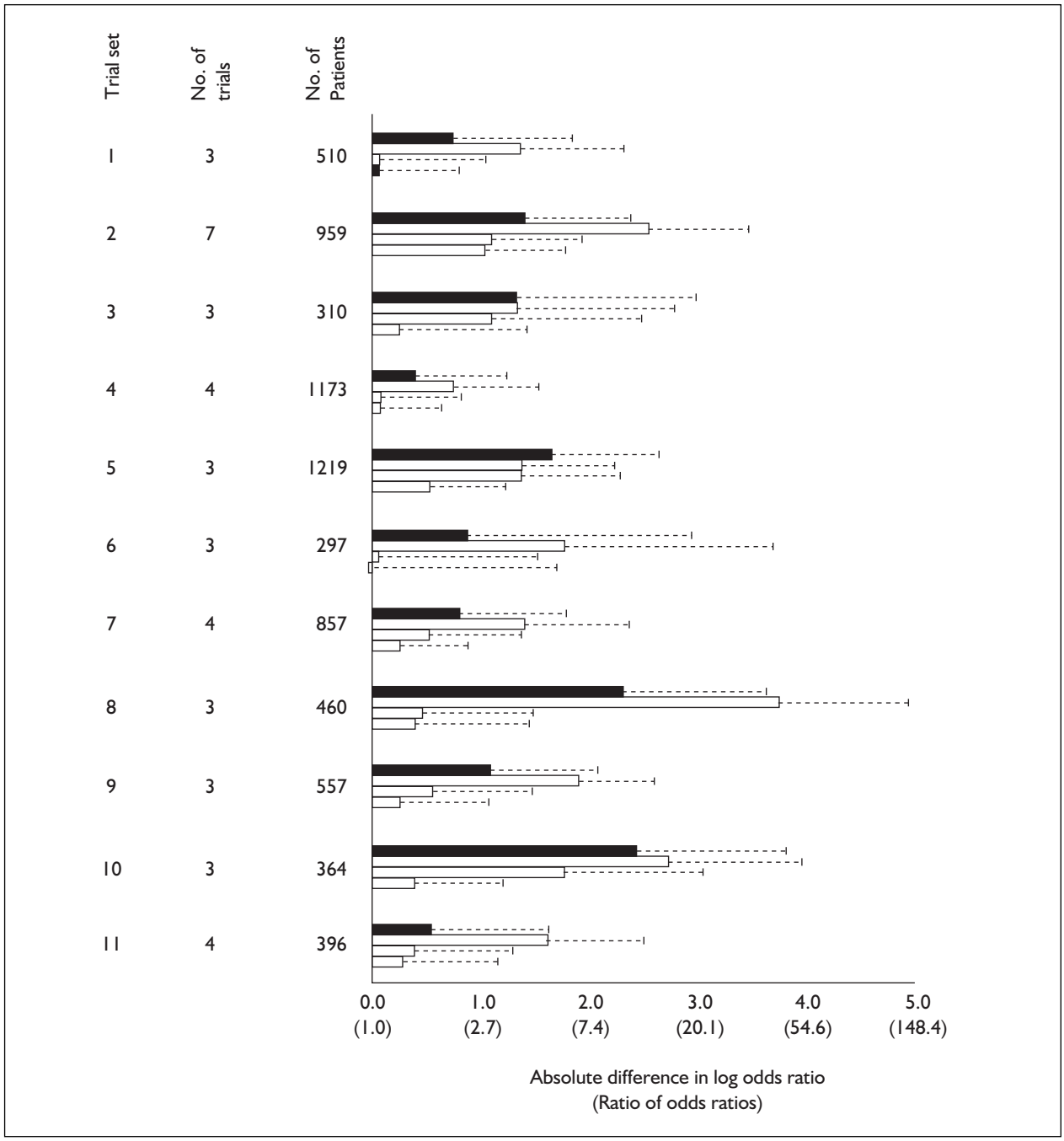

FIGURE 13 *Absolute difference in log odds ratio between the direct and the adjusted indirect comparisons, and between the direct and the naive indirect comparisons. Solid bars, difference between the direct and the adjusted indirect comparison; open bars, difference between the direct and the naive indirect comparison; dotted lines, standard error of difference in log odds ratio. Results are from 11 sets of trials of antimicrobial prophylaxis in colorectal surgery. For each set of trials, the absolute discrepancies by the adjusted indirect comparisons are the same and therefore only one solid bar is presented. Conversely, for each set of trials, absolute discrepancies by three naive indirect comparisons often vary greatly.*

discrepancies ($|z| \geq 1.64$) occurred in five of the 11 sets for the naive indirect comparison (45%) and two of the 11 for the adjusted indirect comparison (18%).

The observed discrepancy between the direct and the indirect comparison may be explained by many possible factors, including chance error, study quality and other factors associated with the internal and external validity of studies. Because of the relatively small sample size (number of trials and patients) in many sets of trials, the discrepancies between the direct and the indirect comparisons may have been exacerbated in this

example. In general, the discrepancy between the direct and the adjusted indirect comparison should be investigated by examining the external and internal validity of the trials involved.

Summary

Five examples from Chapter 6

Although statistically significant discrepancy has been observed between the direct and the adjusted indirect estimate, these examples indicate that such discrepancy may not be clinically important. The direct and the adjusted indirect estimates may be in the same direction and the difference is quantitative rather than qualitative in three[40,107,109] of the five examples. The remaining two examples suggest that the discrepancy between the direct and the adjusted indirect estimate may be clinically important,[27,108] in which statistically significant relative efficacy observed by the direct comparison was not confirmed by the adjusted indirect comparison.

Study-level characteristics reported in the published meta-analyses are not sufficient for detailed investigation of sources of significant discrepancy between the direct and the indirect estimate of relative efficacy of competing interventions.

Example of antimicrobial prophylaxis in colorectal surgery

From a systematic review of antimicrobial prophylaxis in colorectal surgery, 11 sets of randomised trials were identified that could be used to compare antibiotics both directly and indirectly. The discrepancy between the direct and the indirect comparison is defined as the absolute value of difference in log odds ratio.

Considerable discrepancies exist between the direct and the adjusted indirect comparisons. The adjusted indirect comparison has the advantages that the prognostic factors of participants in different trials can be partially taken into account and more uncertainty be incorporated into its result by providing a wider confidence interval. The findings of this section indicate that the discrepancy between the direct and the naive indirect estimate is more unpredictable and more likely to be statistically significant than that between the direct and the adjusted indirect estimate.

Chapter 8

Discussion

Indirect comparisons based on randomised trials: current practice

Indirect comparisons are commonly used for evaluating the relative effectiveness of alternative interventions, with approximately 9.5% of meta-analyses of RCTs identified through DARE including some form of indirect comparison (Chapter 3). Even when only randomised trials are included, the strength of the randomisation procedure is weakened when making indirect comparisons because of the need to combine information across separate sets of trials.

Indirect comparisons are sometimes carried out implicitly and the results interpreted as if from direct comparisons within randomised trials. Indeed, the findings of direct comparisons are sometimes ignored, even when data are available. The use of inferior indirect methods and the inappropriate interpretation of results of indirect comparison may result in misleading estimates of relative efficacy of competing healthcare interventions. Poor methods of analysis could even yield results that are inferior to those obtained through non-randomised studies (as discussed below).

The survey of published meta-analyses indicates that results obtained through indirect comparisons are not always consistent with the findings obtained by direct comparisons. Data available in 28 identified meta-analyses allowed 44 analyses to be undertaken comparing competing interventions both directly and indirectly (Chapter 6). Indirect comparisons can be broadly classified as either naive or adjusted. The naive approach refers to the comparison of results of single arms between different trials. In the adjusted indirect comparison, the comparison of the interventions of interest is adjusted by the results of their direct comparison with a common control group (e.g. placebo). Empirical evidence from the 40 analyses undertaken in Chapter 6 shows the naive approach to be a highly unpredictable method for making indirect comparisons, with a very high frequency of statistically significant discrepancies from the direct estimate. By contrast, for adjusted indirect comparisons there was no clear evidence of significant differences from the direct estimates beyond expectation by chance. However, the amount of direct randomised evidence was rather limited in many of the systematic reviews in the sample.

Performance of different methods of making indirect comparisons

Many systematic reviews were found to include indirect comparisons, yet few reviewers perform formal analyses. Perhaps this lack reflects the fact that there are few publications discussing the analysis of such data (Chapter 4). While the valid methods make assumptions that cannot easily be tested, it may be better to use these explicit formal approaches than for reviewers (and readers) to make such comparisons informally.

The recommended methods can all be described as adjusted indirect comparisons, in which two treatments are compared via their relative effect versus a common comparator. The main types of analysis are the simple weighted combination of separate estimates (e.g. as suggested by Bucher and colleagues[14]), meta-regression and generalised linear models (e.g. logistic regression). Although several examples were found of the use of a naive method in published reviews, no methodological paper with a recommendation to use this strategy was found.

The simulation studies (Chapter 5) showed that all of these methods are unbiased and will give the right answer on average across many such applications. However, the correctness of the standard error of the estimated treatment effect will depend on strong but unverifiable assumptions. Little difference was found in the performance of fixed effect methods for adjusted indirect comparisons, but they tended to give confidence intervals that were too narrow. Partly for this reason, and also in view of additional study heterogeneity compared with the usual case of direct evidence only, a random effects method will usually be appropriate if formal meta-analysis is applied (with the usual provisos regarding the application of such models, such as having enough trials).

The analyses described in Chapter 5 were all based on the data from a single, large trial, with a unified protocol and hence consistent inclusion criteria. Thus, the usual sources of heterogeneity were not present, except for variation in location (country) and some associated case-mix variation. In general, one would expect more variation between independent component studies contributing to an indirect comparison as a result of variation in aspects of the study design. In particular, there may be differences between the two sets of trials being combined with respect to the participants and perhaps also regarding the actual treatments given.

The naive approach, in which the numbers are added as if there was just one trial making each comparison, is problematic even when considering directly randomised trials, and may severely mislead owing to 'Simpson's paradox'.[113] For indirect comparisons, such inappropriate combination is compounded by the discarding of within-trial comparison groups, such that the whole point of randomisation is lost. The results of such analysis are completely untrustworthy, and naive comparisons should never be made.

Quality of evidence from RCTs (direct comparisons)

Evidence from RCTs is, in general, considered to be the best. However, such evidence may be imperfect for several reasons. First, the problem of patient comparability exists in RCTs.[114] The baseline comparability of patients between groups may become problematic owing to a lack of allocation concealment in randomised trials.[115] It is also possible that, purely by chance, the patients randomly allocated to different groups have different prognostic characteristics, particularly when the sample size is small.

Second, after participants have been enrolled there is considerable possibility that biases (such as performance bias, attrition bias and detection bias) may be introduced into trials. For example, patients may drop out after randomisation for various reasons or may be excluded from the analysis. Such withdrawal or exclusion may not be random or balanced across groups. Therefore, even though the patients in different treatment groups are comparable at the time of randomisation, that comparability may disappear owing to non-random exclusion or withdrawal. In addition, lack of blinding in assessment of the outcome may lead to overestimation of treatment effects.[90,115]

Third, empirical evidence has confirmed that published randomised trials may be a biased sample of all trials that have been conducted, owing to publication and related biases.[116] Trials with non-significant or negative results are less likely to be published than trials with significant or positive results. Recently, additional evidence has demonstrated the added problem of selective reporting of outcomes.[117] Therefore, evidence from randomised trials still needs to be interpreted with caution and the possibility of publication bias should be investigated and excluded if possible.

Finally, there may be considerable difference in results among randomised trials, especially small trials. Empirical evidence indicates that subsequent large trials may sometimes overturn conclusions based on small published trials.[118–120]

It has been argued that "randomisation is not sufficient for comparability".[119,121] Here, comparability refers not only to baseline patient characteristics, but also to other study characteristics. It should not be assumed that the adjective 'randomised' is a guarantee of high quality. As a consequence, it is recommended that methodological quality is assessed in a systematic review, although there is no consensus on how to do this or how to make use of the information.[122] Nonetheless, the problems just described apply also to non-RCTs, so while the methodology of directly randomised trials may not always be ideal, such trials do offer the most reliable evidence of treatment efficacy (especially if studies of unacceptably poor quality are discounted).

Thus, when there is a substantial or modest amount of direct evidence the customary practice of ignoring non-randomised studies and indirect randomised evidence will generally be correct. Inevitably, however, in some circumstances there may be few or no trials that have compared directly treatments of particular interest. Thus, the possibility of using indirect evidence becomes important. This report has examined methods for performing indirect comparisons, focusing on indirect evidence from RCTs. The next sections consider issues relating to the use of indirect evidence when direct randomised evidence is either insufficient or non-existent.

Similar problems may apply when performing subgroup analyses within a meta-analysis of trials all making the same direct comparison. If the subgrouping factor relates to the intervention, such as the dose of active treatment, or choice of comparator treatment, then comparison of

subgroups is an indirect comparison of exactly the same type as discussed.

What to do when direct evidence is available but insufficient

Even when some direct randomised evidence is available it is often insufficient. In such cases the adjusted indirect comparison may provide supplementary information in evaluating relative efficacy of competing interventions.[53] Fourteen of the 40 direct comparisons in Chapter 3 are based on one randomised trial, and a median of 18 trials (range 2–96) is incorporated in the corresponding adjusted indirect comparison (Appendix 3). Such large amounts of data available for adjusted indirect comparisons could be used to attempt to strengthen conclusions based on the direct comparisons. Results of the direct and the adjusted indirect comparison can be quantitatively combined to increase statistical power or precision. The statistically non-significant relative effect estimated by the direct comparison may become statistically significant by combining the direct and the adjusted indirect estimate, as happened in three of the 40 comparisons (*Table 11*).

It is also possible that the significant relative effect estimated by the direct comparison becomes non-significant after it has been combined with the adjusted indirect estimate. H_2RA versus sucralfate for non-ulcer dyspepsia from a Cochrane systematic review is an example.[108] Here, the direct comparison based on one randomised trial found that H_2RA was less efficacious than sucralfate (RR 2.74, 95% CI 1.25 to 6.02), while the adjusted indirect comparison indicates that H_2RA was as effective as sucralfate (RR 0.99, 95% CI 0.47 to 2.08). The discrepancy between the direct and the adjusted indirect estimates is marginally statistically significant ($z = 1.84$). The combination of the direct and adjusted indirect estimate provided a statistically non-significant relative risk of 1.56 (95% CI 0.93 to 2.75).

A general question raised by cases like this is: 'when is it appropriate to combine direct and indirect estimates?' Bucher and colleagues[14] concluded that only where direct comparisons are unavailable should indirect comparison meta-analysis be carried out. This is certainly the standard approach, but it is not clear that this will always be the best advice. Even when there is direct randomised evidence, there may be far more information available from indirect comparisons. For example, Song and colleagues[87] examined a case in which one head-to-head randomised trial and six trials contributed to an indirect comparison (via two different comparators). In addition, the use of indirect data may be particularly indicated when it is felt that the methodological quality of the trials making direct comparisons is low.

It will be a matter of judgement whether and how to take into account the indirect evidence. Although a combination of direct and indirect comparisons may appear to strengthen conclusions (by increasing the quantity of data), the increase in precision must be balanced against a loss of confidence in the certainty with which bias is avoided. Few would argue that direct and indirect estimates should always be combined. Rather, many would feel that while presenting separate estimates is necessary, combination will only sometimes be suitable. Some criteria are needed on which to base such a judgement. A very similar problem arises in other contexts within systematic reviews; for example, reviewers ponder whether it is reasonable to combine parallel and cross-over trials, or in epidemiological investigations whether to combine case–control and cohort studies. It is not desirable to base such decisions on whether or not the difference between the two estimates is statistically significant, although this is the easiest approach. A more constructive approach would be to base the decision on the similarity of the participants in the different trials and the comparability of the interventions. Such judgement applies to the directly randomised trials and each subset of trials contributing to the indirect comparison.

What to do when there is no direct evidence

It is not unusual to find that different treatment options have not been directly compared within randomised trials, and conclusions on relative efficacy often end up based entirely on indirect evidence. The adjusted indirect method may be especially useful to obtain some indirect evidence about the relative efficacy of competing interventions. The validity of the indirect estimate was discussed in the previous section. Although the absence of any direct evidence avoids the question of whether to combine, it means that all of the available evidence is indirect. The reliability of that evidence is then of particular concern.

Evidence from an adjusted indirect comparison should be interpreted with caution. The internal

validity of the trials involved in the adjusted indirect comparison should be examined because bias in trials will inevitably infect the results of the adjusted indirect comparison. In addition, for the adjusted indirect estimate to be valid, the key assumption here is that the relative efficacy of an intervention is consistent in patients included in different trials. That is, the estimated relative efficacy should be generalisable. However, generalisability (external validity) of trial results is often questionable because of restricted inclusion criteria, exclusion of patients and the higher level of clinical settings where trials were carried out.[123] Such assessments are more complex still when comparing sets of trials evaluating different treatment comparisons. As in many other situations, interpretation is a matter of judgement and there are no rules applicable across all circumstances.

Are indirect comparisons of RCTs preferable to direct comparisons from non-randomised trials?

As noted above, in the absence of direct randomised evidence one could seek either indirect randomised evidence or non-randomised studies that examine directly the comparison of interest. The authors have not found or generated any empirical evidence to investigate this issue, so this discussion relies on knowledge of the mechanisms that would render both types of comparison biased, and the comparative likelihood that such mechanisms exist.

Non-randomised evaluations of healthcare interventions involve making comparisons between two groups of participants who receive different interventions. If there are any other differences between the groups that are themselves linked to outcome, the comparison will be confounded and potentially biased. For example, the case-mix in the participant groups could differ in terms of age, gender, disease severity or co-morbidities, there could be other differences in treatments between the groups, or the way in which outcomes are defined and assessed could vary. Judgement of the validity of the comparison depends on the degree to which one is assured that 'like is being compared with like' such that one knows that there are no differences between the groups in all factors other than the intervention received. Although many devices are used in non-randomised studies to reduce the potential for such confounding (such as measurement of change scores, matching, stratification and statistical risk adjustment of results), the degree to which these are successful in any individual study is largely inestimable. Empirical studies have found, however, that adjustment may fail to remove the bulk of selection bias.[15] In addition, as there is always a likelihood of prognostic factors that are unknown or unmeasurable, adjustment methods cannot cope with all eventualities. Indirect comparisons also feature comparisons between non-randomised groups; however, in this instance the groups are not cohorts receiving one or other treatment, but randomised trials within which different comparisons are made. Now the same differences in case-mix, concomitant therapy and follow-up could exist between the trials in an indirect comparison in the same way as they do between the groups within a non-randomised study. However, even if the magnitudes of these differences are similar (and there are reasons to argue that they are most likely to be less in indirect comparisons), there are reasons to believe that the bias they introduce would be less than in a non-randomised study.

For binary data the mechanism by which this prediction is reached is as follows. Variations in potentially confounding factors will change the average event rate observed in the trial as they do in a non-randomised study, as they relate to the frequency of outcome. However, it is likely that their impact on outcome would affect both groups in each trial in a proportionate manner. If this happens, and if the treatment effect is calculated using an appropriate metric (probably a risk ratio or an odds ratio), very little variation in the treatment effect will be observed between trials of the same comparison among participants with varying baseline risks. When this is the case the indirect comparison will be unbiased. This observation is fundamental to the practice of meta-analysis.[57]

The case of continuous outcomes seems not to have been considered in the same light. Although the broad concerns are the same, the impact of effect modifiers, including case-mix variation, is unpredictable; it may work in additive or multiplicative manners, affecting either or both of the mean and standard deviation.

An indirect comparison will, however, be biased if the differences in baseline risk between trials are linked to differences in the observed treatment effect. This would occur if any of the factors varying between the trials are known effect modifiers (or, in meta-analysis terminology, sources

of heterogeneity). Thus, it would be wrong to claim, based on the argument of constancy of treatment effects, that indirect comparisons are always unbiased. There are examples where such effect modifiers exist such that indirect comparisons will be biased. Therefore, it will always be desirable to show similar distributions of confounding factors in the trials included in an indirect comparison, rather than rely on the assumption of constancy of effect across varying baseline risk.

The resampled results (Chapter 5), despite demonstrating heterogeneity between trials, probably underestimate the potential for such bias, as all the trials in the reconstructed analysis used exactly the same protocol. Such uniformity of protocol is unlikely in reality.

In conclusion, bias is less likely in indirect comparisons than within a non-randomised study, as an indirect comparison between treatment effects estimated in trials with different baseline risks is not necessarily biased, whereas the same difference in baseline risks between groups in a non-randomised comparison is sufficient to render it biased. The present authors thus agree with the relative placing of these levels of evidence by McAlister and colleagues,[5] as discussed in Chapter 1.

Comparing multiple interventions simultaneously

The focus of this review and empirical research has been the use of indirect comparisons to compare two prespecified healthcare interventions. As noted in Chapter 1, reviewers sometimes consider simultaneously several interventions with the natural desire to say something about their relative efficacy. The results of such a review may be presented in the form of a league table of efficacy.

This situation was not studied in the present empirical work, although it did consider indirect comparison of two specific treatments from trials involving four different treatments (Chapter 5). Here, some different situations within this broad framework are briefly summarised, and an indication is given of how analysis might proceed. Distinction is made between two rather different contexts in which multiple treatments are considered simultaneously.

First, there may be several competing interventions, each of which has been compared against the same comparator in one or more RCTs. The most obvious example is where each of several drugs has been evaluated in placebo-controlled trials. There are several such examples in pain research.[124,125] The common comparator makes the comparison of the various drugs seem relatively simple. In the particular case of trials of pain relief there is also much greater consistency of trial methodology (notably standardised outcome measure) than is seen in most medical areas. However, the quoted results are often simply separate meta-analyses of each active drug versus placebo, ranked by treatment effect (perhaps NNT); this display does not allow an easy assessment of the difference between any two particular treatments, and so may give a false impression of their relative merits. Also, when data sets are presented in this way, there may be a suspicion that any direct comparisons have been omitted to simplify the presentation.

Sometimes researchers making such comparisons use the naive approach by summarising treatment arms across trials,[6] or ignore the control groups altogether.[46] The strong warnings given against the naive approach apply even more strongly in the multiple treatment case.

This data structure is similar to that where there are several trials of the same intervention versus control, but where that intervention was delivered in different ways in different trials; for example, dose, regimen or route may vary across trials. However, the aim of meta-analysis in such cases is usually to evaluate the treatment against the comparator rather than to study explicitly the importance of the mode of delivery.

The second case is where there are multiple studies that have each compared two or more of a set of treatments, but without a common comparator. This more general situation is exemplified by the trial of antithrombotic therapy considered in Chapter 4 and also a set of trials of thrombolytic therapy discussed by Hasselblad and Kong,[68] which is shown in *Table 17*.

The aim here may be to estimate the effect of a particular drug versus placebo when no such trial has been done, but such data may also be the basis of an attempt to rank the treatments. Data such as these can arise when new treatments are introduced over a period of several years, an effect seen more clearly in the subset of trials shown in *Table 18*. This subset perhaps makes it clearer that estimation is based on a chain of inference. For example, adjusted indirect comparisons or logistic

TABLE 17 *Treatments studied in 17 randomised trials reporting 30-day mortality examining thrombolytic therapy for patients with acute coronary syndrome (treatment within 6 hours of onset) (see Hasselblad and Kong[68] for further details)*

Trial	Placebo	t-PA	APSAC	SK	RPA	Accelerated t-PA
ASSENT	✓	✓				
ECGS	✓	✓				
AIMS	✓		✓			
Bassand-1	✓		✓			
GISSI	✓			✓		
ISIS-2	✓			✓		
ISAM	✓			✓		
TEAM-3		✓	✓			
Bassand-2		✓	✓			
GISSI-2		✓		✓		
ISG		✓		✓		
ISIS-3		✓	✓	✓		
TIMI-4			✓		✓	
TAPS			✓			✓
INJECT				✓	✓	
GUSTO I				✓		✓
GUSTO III					✓	✓

APSAC, anisoylated plasminogen streptokinase activator complex; RPA, reteplase; SK, streptokinase; tPA, tissue plasminogen activator.

TABLE 18 *Subset of trials from* Table 17*, showing sequential evaluation of new treatments*

Trial	Placebo	t-PA	APSAC	SK	RPA	Accelerated t-PA
ASSENT	✓	✓				
TEAM-3		✓	✓			
ISIS-3			✓	✓		
INJECT				✓	✓	
GUSTO III					✓	✓

APSAC, anisoylated plasminogen streptokinase activator complex; RPA, reteplase; SK, streptokinase; tPA, tissue plasminogen activator.

regression could be used to estimate the effect of accelerated t-PA versus placebo. Such an analysis would link the results of five trials, three of which did not investigate either of the treatments of interest. Like all chains, its strength depends on the weakest link. Here TEAM-3 was a very small trial of 325 patients, with an imprecise estimate of treatment effect: it would therefore be unsatisfactory to rely on this chain of inference. ISIS-3 enrolled over 17,000 patients to make the same comparison, so there would be no need to rely on TEAM-3. The amount of information in each component of such an inferential chain would certainly be a concern, even though the uncertainty of the final estimate would take sample size into account. Other concerns would include the methodological quality and the degree of comparability of the treatments, participants and protocols of the component trials. Hasselblad and Kong rather underplay the assumptions behind such an analysis. They suggest that the method is acceptable as long as there is a common subpopulation of participants in the different trials.[68] In fact, the more links there are the stronger will be the assumptions of adjusted incorrect comparisons. Recently, other authors have proposed new models for analysing such networks of comparisons.[76,88]

Chapter 9
Conclusions

Implications for practice of systematic reviews

When conducting systematic reviews to evaluate the effectiveness of interventions, direct evidence from good-quality RCTs should be used wherever possible. If no such evidence exists a call for further trials may be necessary (if they are deemed ethical). If further research is not feasible, it may be necessary to look for direct comparisons in non-randomised studies and/or indirect comparisons from RCTs, which would require additional searches of the literature. The reviewer needs, however, to be aware that both are susceptible to bias, although bias is less likely in indirect comparisons. The use of non-randomised studies within a systematic review is, in addition, perhaps more problematic in terms of identifying the studies. The development of better indexing and sensitive search strategies is required to aid the identification of study designs other than RCTs. It should also be noted, however, that the development of search strategies for the identification of RCTs may actually exclude data that could be used when making indirect comparisons, particularly if a search is drug or treatment specific. Ideally, indirect comparisons should be prespecified in a review's protocol and the search strategy developed so as to include all relevant drug/treatment comparisons.

When making indirect comparisons in a systematic review, the reviewer needs to be clear that that is what they are doing, and interpret the results appropriately. The naive approach, comparing the results of single arms between trials, should be avoided as empirical evidence shows it to have a high frequency of statistically significant discrepancies from the direct estimate. Ideally, an adjusted indirect comparison method should be used, using the random effects model (given the increase in study heterogeneity). The main types of adjusted indirect comparison, showing little difference in performance, are the adjusted indirect comparison, meta-regression and logistic regression.

If both direct and indirect comparisons are possible within a review, these should be done separately before considering whether to pool data. Whether or not the two sets of information are combined, or when only indirect comparisons are available, interpretations should be made even more cautiously than in a standard meta-analysis of just head-to-head randomised trials, in view of the partially observational nature of the comparisons.

Implications for clinical research

Indirect comparisons can be used to overcome certain problems that may arise in the design and conduct of an RCT. An example can be seen in a proposed European multicentre trial of surgical techniques for repairing cleft lip and palate.[126] Each centre approached used a different surgical technique and was only willing to participate in the trial if their 'preferred' surgical technique was one of the treatment arms to which patients would be randomised. To overcome this, a common protocol was devised that allowed each centre to randomise to one of two surgical techniques: their usual, preferred technique, and an alternative technique that was common to all participating centres. The alternative technique was planned to be used as the common comparator, allowing adjusted indirect comparisons to be made between the centres' preferred surgical techniques. A head-to-head comparison of the different surgical techniques may have been preferable, but the indirect comparison was able to provide an acceptable design for the surgeons, who otherwise may not have been willing to participate in the trial.

When considering the use of an indirect comparison when designing such a multicentre RCT, consideration should be given to devising a unified protocol with consistent inclusion criteria across all centres.

Recommendations for methodological research

The majority of the indirect comparisons identified were made within meta-analyses of binary data. There is a need for evaluation of alternative methods for analysis of indirect comparisons for continuous data.

In addition, there is a need for empirical research into how different methods of indirect comparison perform in cases where there is a large treatment effect.

Further research is required to consider how to determine when it is appropriate to look at indirect comparisons and how to judge when to combine both direct and indirect comparisons. Research into how evidence from indirect comparisons compares to evidence from non-randomised studies may also be warranted. (This question has not been considered in the current project.)

The empirical investigations (Chapter 5) were based on one large, multicentre trial[101] with a common protocol across each centre. It would be useful to repeat the investigations using individual patient data from a meta-analysis of several RCTs using different protocols.

The odds ratio was used as the measure of effect within the simulation study. Although logistic regression calls for the effect measure to be the odds ratio, it would be interesting to evaluate the impact of choosing different binary effect measures for the inverse variance method.

Acknowledgements

The authors would like to thank Kath Wright (CRD, York) for her assistance with the development of the electronic search strategies, and members of the International Stroke Trial Collaborative Group for allowing data from their multicentre RCT to be used for the empirical investigations within this report. We would also like to thank the referees who provided valuable comments on the draft report.

Contribution of authors

DG Altman (Professor of Statistics in Medicine), JJ Deeks (Senior Medical Statistician), AJ Eastwood (Senior Research Fellow), AM Glenny (Lecturer in Evidence Based Oral Health Care) and F Song (Reader in Research Synthesis) developed the structure of the report.

AJE and AMG coordinated the project.

In developing the search strategy, the initial searches were developed and undertaken by Kath Wright (Centre for Reviews and Dissemination, University of York) and the update searches were undertaken by AMG.

Handsearching was carried by DGA (Chapter 4), AJE (Chapter 3), AMG (Chapter 3), and FS (Chapters 3 and 6).

DGA, JJD, AJE, AMG and FS carried out the screening of search results and retrieved papers against inclusion criteria.

The analysis of data was carried out by DGA (Chapters 4 and 5), M Bradburn (Statistician) (Chapter 5), R D'Amico (Research Fellow) (Chapter 5), JJD (Chapters 4 and 5), AJE (Chapter 3), AMG (Chapter 3), C Sakarovitch (Biostatistician) (Chapter 5) and FS (Chapters 3, 6 and 7).

Production of the full report was carried out by DGA, JJD, AJE, AMG and FS.

References

1. Pocock SJ. *Clinical trials: a practical approach*. New York: John Wiley; 1996.
2. Moher D, Schulz KF, Altman D. CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;**285**:1987–91.
3. Song F, Glenny AM. Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials. *Health Technol Assess* 1998;**2**(7).
4. de Craen AJM, Tijssen JGP, de Gans J, Kleijnen J. Placebo effect in the acute treatment of migraine: subcutaneous placebos are better than oral placebos. *J Neurol* 2000;**247**:183–8.
5. McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999;**282**:1371–7.
6. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses. *Arthritis Rheum* 1990;**33**:1449–61.
7. Garg R, Yusuf S, for the Collaborative Group on ACE Inhibitor Trials. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. *JAMA* 1995;**273**:1450–6.
8. Homocysteine Lowering Trialists' Collaboration. Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomised trials. *BMJ* 1998;**316**:894–8.
9. Boersma E, Maas AC, Deckers JW, Simoons ML. Early thrombolytic treatment in acute myocardial infarction: reappraisal of the golden hour. *Lancet* 1996;**348**:771–5.
10. Moore A, McQuay H, Gray M, editors. *Bandolier*. Oxford: Pain Relief Unit. URL: http://www.jr2.ox.ac.uk/bandolier/
11. Tramer MR, Moore RA, McQuay HJ. Prevention of vomiting after paediatric strabismus surgery: a systematic review using the numbers-needed-to-treat method. *Br J Anaesth* 1995;**75**:556–61.
12. de Craen AJM, Vickers AJ, Tijssen JGP, Kleijnen J. Number-needed-to treat and placebo-controlled trials. *Lancet* 1998;**351**:310.
13. Murray DW, Britton A, Bulstrode CJ. Thromboprophylaxis and death after total hip replacement. *J Bone Joint Surg Am* 1996; **78-B**:863–70.
14. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;**50**:683–91.
15. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al.* Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**(27).
16. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. II. Maintenance of vascular graft or arterial patency by antiplatelet therapy. *BMJ* 1994; **308**:159–68.
17. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994;**308**:81–106.
18. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. III. Reduction in venous thrombosis and pulmonary embolism by antiplatelet prophylaxis among surgical and medical patients. *BMJ* 1994; **308**:235–46.
19. Lowenthal A, Buyse M. Secondary prevention of stroke: does dipyridamole add to aspirin? *Acta Neurol Belg* 1994;**94**:24–34.
20. Li Wan Po A, Zhang WY. Systematic overview of co-proxamol to assess analgesic effects of addition of dextropropoxyphene to paracetamol. *BMJ* 1997;**315**:1565–71.
21. Marshall JK, Irvine EJ. Rectal corticosteroids versus alternative treatments in ulcerative colitis: a meta-analysis. *Gut* 1997;**40**:775–81.
22. Matchar DB, McCrory DC, Barnett HJ, Feussner MD. Medical treatment for stroke prevention. *Ann Intern Med* 1994;**121**:41–53.
23. Milne RJ, Olney RW, Gamble GD, Turnidge J. Tolerability of roxithromycin vs erythromycin in comparative clinical trials in patients with lower respiratory tract infections. *Clinical Drug Investigation* 1997;**14**:405–17.
24. Moore A, Collins S, Carroll D, McQuay H. Paracetamol with and without codeine in acute pain: a quantitative systematic review. *Pain* 1997;**70**:193–201.

25. Pope JE, Anderson JJ, Felson DT. A meta-analysis of the effects of nonsteroidal anti-inflammatory drugs on blood pressure. *Arch Intern Med* 1993; **153**:477–84.

26. Poynard T, Leroy V, Cohard M, Thevenot T, Mathurin P, Opolon P, *et al.* Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: effects of dose and duration. *Hepatology* 1996;**24**:778–89.

27. Zhang WY, Li Wan Po A. Analgesic efficacy of paracetamol and its combination with codeine and caffeine in surgical pain – a meta-analysis. *J Clin Pharm Ther* 1996;**21**:261–82.

28. Piccinelli M, Pini S, Bellantuono C, Wilkinson G. Efficacy of drug treatment in obsessive–compulsive disorder: a meta-analytic review. *Br J Psychiatry* 1995;**166**:424–43.

29. Zalcberg JR, Siderov J, Simes J. The role of 5-fluorouracil dose in the adjuvant therapy of colorectal cancer. *Ann Oncol* 1996;**7**:42–6.

30. Leizorovicz A. Comparison of the efficacy and safety of low molecular weight heparins and unfractionated heparin in the initial treatment of deep venous thrombosis. An updated meta-analysis. *Drugs* 1996;**7**:30–7.

31. Pignon JP, Arriagada R, Ihde DC, Johnson DH, Perry MC, Souhami RL, *et al*. A meta-analysis of thoracic radiotherapy for small-cell lung cancer. *N Engl J Med* 1992;**327**:1618–24.

32. Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer; a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ* 1995;**311**:899–909.

33. Rossouw JE. Lipid-lowering interventions in angiographic trials. *Am J Cardiol* 1995;**76**:86–92C.

34. Koch M, Dezi A, Ferrario F, Capurso L. Prevention of nonsteroidal anti-inflammatory drug-induced gastrointestinal mucosal injury. *Arch Intern Med* 1996;**156**:2321–32.

35. Holme I. Cholesterol reduction and its impact on coronary artery disease and total mortality. *Am J Cardiol* 1995;**76**:10–17C.

36. Moore RA, McQuay HJ. Single-patient data meta-analysis of 3453 postoperative patients: oral tramadol versus placebo, codeine and combination analgesics. *Pain* 1997;**69**:287–94.

37. Aro J. Cardiovascular and all-cause mortality in prostatic cancer patients treated with estrogens or orchiectomy as compared to the standard population. *Prostate* 1991;**18**:131–7.

38. Poynard T, Lemaire M, Agostini H. Meta-analysis of randomized clinical trials comparing lansoprazole with ranitidine or famotidine in the treatment of acute duodenal ulcer. *Eur J Gastroenterol Hepatol* 1995;**7**:661–5.

39. Lefering R, Neugebauer EAM. Steroid controversy in sepsis and septic shock: a meta-analysis. *Crit Care Med* 1995;**23**:1294–303.

40. Chiba N, de-Gara CJ, Wilkinson FM, Hunt RH. Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: a meta-analysis. *Gastroenterology* 1997;**112**:1798–810.

41. Coulter A, Kelland J, Peto V, Rees MC. Treating menorrhagia in primary care: an overview of drug trials and a survey of prescribing practice. *Int J Technol Assess Health Care* 1995;**11**:456–71.

42. Bansal VK, Beto JA. Treatment of lupus nephritis: a meta-analysis of clinical trials. *Am J Kidney Dis* 1997;**29**:193–9.

43. Schmieder RE, Martus P, Klingbeil A. Reversal of left ventricular hypertrophy in essential hypertension. A meta-analysis of randomized double-blind studies. *JAMA* 1996;**275**:1507–13.

44. Unge P, Berstad A. Pooled analysis of anti-*Helicobacter pylori* treatment regimens. *Scand J Gastroenterol Suppl* 1996;**220**:27–40.

45. Imperiale TF, Speroff TS. A meta-analysis of methods to prevent venous thromboembolism following total hip replacement. *JAMA* 1994; **271**:1780–5.

46. Conlin PR, Spence JD, Williams B, Ribeiro AB, Saito I, Benedict C, *et al*. Angiotensin II antagonists for hypertension: are there differences in efficacy? *Am J Hypertens* 2000;**13**:418–26.

47. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. *Lancet* 1994;**343**:311–22.

48. Ekbom A, Taube A. Methodological flaws in study on prostatic cancer. *Prostate* 1994;**25**:53–4.

49. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness. CRD guidelines for those carrying out or commissioning reviews*. Report No. 4. York: NHS Centre for Reviews and Dissemination; 1996.

50. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Stat Med* 1996;**15**:537–57.

51. Dominici F, Parmigiani G, Wolpert RL, Hasselblad V. Meta-analysis of migraine headache treatments: combining information from heterogeneous designs. *Journal of the American Statistical Association* 1999;**94**:16–28.

52. Hasselblad V. Meta-analysis of multitreatment studies. *Med Decis Making* 1998;**18**:37–43.

53. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;**15**:2733–49.

54. Hirotsu C, Yamada Y. Estimating odds ratios through the connected comparative experiments. *Communications in Statistics – Theory and Methods* 1999;**28**:905–29.

55. Gleser LJ, Olkin I. Meta-analysis for 2 × 2 tables with multiple treatment groups. In Stangl DK, Berry DA, editors. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker; 2001. pp. 179–89.

56. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In Cooper H, Hedges LV, editors. *The handbook of research synthesis*. New York: Russell Sage; 1994. pp. 340–55.

57. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;**21**:1575–600.

58. Davey Smith G, Egger M. Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials. In Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in healthcare. Meta-analysis in context*. 2nd ed. London: BMJ Books; 2001. pp. 143–56.

59. Thompson SG. Why and how sources of heterogeneity should be investigated. In Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in healthcare. Meta-analysis in context*. 2nd ed. London: BMJ Books; 2001. pp. 157–75.

60. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Books; 2001. pp. 285–312.

61. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003; **326**:219.

62. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;**21**:1559–73.

63. Anderson S, Hauck WW. The transitivity of bioequivalence testing: potential for drift. *Int J Clin Pharmacol Ther* 1996;**34**:369–74.

64. Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. *Drug Information Journal* 1999;**33**:109–18.

65. Mainland D. *The treatment of clinical and laboratory data*. Edinburgh: Oliver and Boyd; 1938. pp. 151–3.

66. Armitage P, Berry G. The sampling error of a difference. In *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell; 1987. pp. 88–90.

67. Eddy DM, Hasselblad V, Schachter R. *Meta-analysis by the confidence profile method*. Boston, MA: Academic Press; 1992.

68. Hasselblad V, Kong DF. Statistical methods for comparison to placebo in active–control trials. *Drug Information Journal* 2001;**35**:435–49.

69. Packer M, Antonopoulos GV, Berlin JA, Chittams J, Konstam MA, Udelson JE. Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: results of a meta-analysis. *Am Heart J* 2001;**141**:899–907.

70. Fisher LD, Gent M, Butler HR. Active–control trials: how would a new agent compare with placebo? A method illustrated with clopidogrel, aspirin, and placebo. *Am Heart J* 2001;**141**:26–32.

71. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol* 2002;**2**:13.

72. Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analysis? *BMJ* 1994;**309**:1007–10.

73. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001;**20**:2219–41.

74. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes *Stat Med* 2005;**24**:1307–19.

75. Brown H, Prescott R. *Applied mixed models in medicine*. Chichester: John Wiley; 1999.

76. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med* 2003;**22**:2995–3016.

77. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: John Wiley; 2000.

78. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;**18**:2693–708.

79. Lim E, Ali A, Routledge T, Edmonds L, Altman DG, Large S. Indirect comparison meta-analysis of aspirin therapy after coronary surgery. *BMJ* 2003; **327**:1309–11.

80. Lim E. First principles or evidence based critique? URL: http://bmj.bmjjournals.com/cgi/eletters/327/7427/1309#46233 (accessed September 2004).

81. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. Chichester: John Wiley; 2004.

82. Phillips A. Trial and error: cross-trial comparisons of antiretroviral regimens. *AIDS* 2003;**17**:619–23.

83. Senn S. Active control equivalence trials. *Statistical issues in drug development*. Chichester: John Wiley; 2002. pp. 207–17.

84. Siegel JP. Equivalence and non-inferiority trials. *Am Heart J* 2000;**139**:S166–70.

85. ICH Harmonised Tripartite Guideline. Choice of control group and related issues in clinical trials. URL: http://www.nihs.go.jp/dig/ich/efficacy/e10/e10step4.pdf (accessed March 2005).

86. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000;**19**:1707–28.

87. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy – illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000;**21**:488–97.

88. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;**21**:2313–24.

89. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;**365**:1159–62.

90. Lambert PC, Sutton AJ, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002;**55**:86–94.

91. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data meta-regression for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;**21**:371–87.

92. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2000;**19**:3417–32.

93. Begg CB, Pilote L. A model for incorporating historical controls into a meta-analysis. *Biometrics* 1991;**47**:899–906.

94. Li Z, Begg CB. Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *Journal of the American Statistical Association* 1994;**89**:1523–7.

95. Raghunathan TE. Pooling controls from different studies. *Stat Med* 1991;**10**:1417–26.

96. Büchner T, Dohner H, Ehninger G, Ganser A, Hasford J. Up-front randomization and common standard arm: a proposal for comparing AML treatment strategies between different studies. *Leuk Res* 2002;**26**:1073–5.

97. Büchner T, Dohner H, Ehninger G, Ganser A, Niederwieser D, Hasford J. Cross-trial networking in AML: a step forward rather than corner cutting. *Leuk Res* 2004;**28**:649–50.

98. Hills RK, Richards SM, Wheatley K. Corner cutting compromises clinical trials: the inherent problems with up-front randomisation and a common standard arm. *Leuk Res* 2003;**27**:1071–3.

99. Tsong Y, Wang SJ, Hung HM, Cui L. Statistical issues on objective, design, and analysis of noninferiority active-controlled clinical trial. *J Biopharm Stat* 2003;**13**:29–41.

100. Wang SJ, Hung HM, Tsong Y. Utility and pitfalls of some statistical methods in active controlled clinical trials. *Control Clin Trials* 2002;**23**:15–28.

101. Sandercock PAG, Collins R, Counsell C, Farrell B, Peto R, Slattery J, *et al.* The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *Lancet* 1997;**349**:1569–81.

102. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in Stata™. In Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Books; 2001. pp. 347–85.

103. Bradburn MJ, Deeks JJ, Altman DG. Metan-an alternative meta-analysis command (sbe24) [correction appears in *Stata Technical Bulletin*, No. 45, 21 (sbe24.1)]. *Stata Technical Bulletin* 1998; **44**:4–15.

104. Sharp S. sbe23: Meta-analysis regression. *Stata Technical Bulletin* 1998;**42**:16–24.

105. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic-regression models. *International Statistical Review* 1991; **59**:227–40.

106. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;**326**:472–5.

107. Rostom A, Wells G, Tugwell P, Welch V, Dube C, McGowan J. Prevention of chronic NSAID induced upper gastrointestinal toxicity. In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2000.

108. Soo S, Moayyedi P, Deeks J, Delaney B, Innes M, Forman D. Pharmacological interventions for non-ulcer dyspepsia. In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2000.

109. van Pinxteren B, Numans ME, Bonis PA, Lau J. Short-term treatment with proton pump inhibitors, H_2-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopy negative reflux disease. In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2000.

110. Palmer BV, Mannur KR, Ross WB. An observer blind trial of co-amoxiclav versus cefuroxime plus metronidazole in the prevention of postoperative wound infection after general surgery. *J Hosp Infect* 1994;**26**:287–92.

111. Rowe-Jones DC, Peel ALG, Kingston RD, Shaw JFL, Teasdale C, Cole DS. Single dose cefotaxime plus metronidazole versus three dose cefuroxime plus

metronidazole as prophylaxis against wound infection in colorectal surgery: multicentre prospective randomised study. *BMJ* 1990; **300**:18–22.

112. Kwok SPY, Lau WY, Leung KL, Ku KW, Ho WS, Li AKC. Amoxycillin and clavulanic acid versus cefotaxime and metronidazole as antibiotic prophylaxis in elective colorectal resectional surgery. *Chemotherapy* 1993;**39**:135–9.

113. Altman DG, Deeks JJ. Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC Med Res Methodol* 2002;**2**:3.

114. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; **335**:149–53.

115. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**:408–12.

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

117. Chan A-W, Hròbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;**291**:2457–65.

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

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

120. 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.

121. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;**52**:487–97.

122. Jüni P, Altman DG, Egger M. Assessing the quality of randomised controlled trials. In Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care. Meta-analysis in context*. London: BMJ Books; 2001. pp. 87–108.

123. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215–18.

124. McQuay HJ, Moore RA. *An evidence-based resource for pain relief*. Oxford: Oxford University Press; 1998.

125. Bandolier. The Oxford league table of analgesic efficacy. URL: http:// www.jr2.ox.ac.uk/bandolier/booth/painpag/acutrev/analgesics/lftab.html (accessed September 2004).

126. European Collaboration on Craniofacial Anomalies (Eurocran). URL: http://www.eurocran.org/content.asp?contentID=1345 (accessed September 2004).

127. Sackett DL. Rules of evidence and clinical recommendations on the use of asymptomatic atherosclerosis. *Stroke* 1989;**20**:844–9.

128. Poynard T. Evaluation de la qualite methodolgique des essais therapeutiques randomises. *Presse Med* 1988;**17**:315–88.

129 Marshall JK, Irvine EJ. Rectal aminosalicylate therapy for distal ulcerative colitis: a meta-analysis. *Aliment Pharmacol Ther* 1995;**9**:293–300.

130. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D *et al*. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;**2**:31–49.

131. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; **4**:942–51.

132. Evans M, Pollock AV. A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection. *Br J Surg* 1985;**72**:265–70.

133. Jadad AR, Moore RA, Carrol D, Jenkinson C, Reynolds DJM, Gavaghan DJ, *et al*. Assessing the quality of reports of randomised clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**:1–12.

134. Poynard T, Pignon JP. *Duodenal ulcer. Analysis of 293 randomized clinical trials*. London: Montrouge, John Libbey Eurotext; 1989.

135. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;**306**:1332–7.

136. Schmieder RE, Schlaich MP, Klingbeil AU, Martus P. Update on reversal of left ventricular hypertrophy in essential hypertension (a meta-analysis of all randomized double-blind studies until December 1996). *Nephrol Dial Transplant* 1998;**13**:564–9.

137. Cheng L, Gulmezoglu AM, Ezcurra E, van-Look PFA. Interventions for emergency contraception (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2000.

138. Collins SL, Moore RA, McQuay HJ, Wiffen PJ, Edwards JE. Single dose oral ibuprofen and diclofenac for postoperative pain (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2000.

139. Delaney BC, Innes MA, Deeks JJ, Wilson S, Oakes R, Moayyedi P, *et al*. Initial management strategies for dyspepsia (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001

140. Di Mario F, Battaglia G, Leandro G, Grasso G, Vianello F, Vigneri S. Short-term treatment of gastric ulcer. A meta-analytical evaluation of blind trials. *Dig Dis Sci* 1996;**41**:1108–31.

141. Handoll HHG, Farrar MJ, McBirnie J, Tytherleigh-Strong G, Awal KA, Milne AA, *et al*. Heparin, low molecular weight heparin and physical methods for preventing deep vein thrombosis and pulmonary embolism following surgery for hip fractures (Cochrane Review). In *The Cochrane Library* (Issue 4). Oxford: Update Software; 2002.

142. Horn J, Limburg M. Calcium antagonists for acute ischemic stroke (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001.

143. McIntosh HM, Olliaro P. Artemisinin derivatives for treating uncomplicated malaria (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001.

144. Silagy C. Physician advice for smoking cessation (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001.

145. Silagy C, Mant D, Fowler G, Lancaster T. Nicotine replacement therapy for smoking cessation (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001.

146. Trindade E, Menon D. Selective serotonin reuptake inhibitors (SSRIs) for major depression. Part I. Evaluation of the clinical literature. Report No. 3E. Ottawa, Ontario: Canadian Coordinating Office for Health Technology Assessment; 1997 August 1997.

147. Zhang WY, Li Wan Po A. Efficacy of minor analgesics in primary dysmenorrhoea: a systematic review. *Br J Obstet Gynaecol* 1998;**105**:780–9.

148. Ausejo M, Saenz A, Pham B, Kellner JD, Johnson DW, Moher D, *et al*. Glucocorticoids for croup (Cochrane Review). In *The Cochrane Library* (Issue 3). Oxford: Update Software; 2001.

149. Sauriol L, Laporta M, Edwardes MD, Deslandes M, Ricard N, Suissa S. Meta-analysis comparing newer antipsychotic drugs for the treatment of schizophrenia: evaluating the indirect approach. *Clin Ther* 2001;**23**:942–56.

150. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med* 1999;**131**:492–501.

151. Singer DE. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation. *Ann Intern Med* 2000;**132**:841–2.

152. EAFT (European Atrial Fibrillation Trial) Study Group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. *Lancet* 1993;**342**:1255–62.

Appendix 1

Reviews of effectiveness prescreening form

Author/year DARE Accession number	Relevant project (E/I/N)	Useful (✓/?/×)			Type of study design			M-A	ENRIS		IC			Comments	Methodology citations
		E	I	B	RCTs	Comp	Non-comp		VA	CF	IC done	DC/IC done	Suff. data		

Key:

1. Author/year and DARE accession number if relevant.
2. Relevant project
   - E: ENRIS;
   - I: indirect comparison
   - N: neither
   - Note: can be both E and I.
3. Useful
   - ✓: if adequate data are presented for use in either review
   - ?: if unsure or insufficient data are presented, but there is a possibility of obtaining further data, or if meta-analysis of non-RCTs but no validity assessment or control for confounding variables has been undertaken (specifically for ENRIS)
   - X: if there are insufficient data for use in either review (mark relevant column)
   - B: mark if the paper is useful but only for the background/discussion.
4. Included study design — Enter number of RCTs, comparative and non-comparative studies.
5. M-A
   - ✓: if meta-analysis has been undertaken
   - X: if no meta-analysis has been undertaken (narrative review).
6. ENRIS
   - VA: need to know whether some form of validity assessment has been undertaken
   - CF: need to know whether they have controlled for confounding variables
7. IC
   - IC done: has an indirect comparison been conducted in the review?
   - IC/DC done: does the review present both direct and indirect comparisons of data?
   - Suff. data: are there sufficient primary data (i.e. number of events and sample size for each arm of the included trials) or summary statistics and SE/CIs, in order to force an indirect comparison, where a direct comparison has been done? If so, state comparisons for which an IC is possible in the comments field.
8. Comments
9. Methodology citations — List citations of potentially relevant methodology papers cited in the review (reference lists of all reviews should be examined).

NB. All columns should be completed.

Appendix 2

Data extraction form

Author (year)			
Title			
Journal			
Objectives			
COMPARISONS BEING MADE FOR RCTs			
INDIRECT		Done ✓	Not Done*
Intervention A / ctrl	Intervention B / ctrl	No. of trials A / B	
DIRECT (only those that are relevant to ICs)		Done ✓	Not Done
Intervention A	Intervention B	No. of trials	
METHODS USED FOR CONDUCTING IC			

RESULTS OF IC		
Intervention A / ctrl	Intervention B /ctrl	Summary statistic
RESULTS OF DC		
Intervention A	Intervention B	Summary statistic
INTERPRETATION OF RESULTS IC interpreted appropriately?		Yes ☐ No ☐ Unsure ☐
How were ICs interpreted? Details:		
Other COMMENTS		

*But sufficient data presented in paper to force IC.

Appendix 3

Table of systematic reviews making indirect comparisons

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
ATC[17] **Validity assessment undertaken within the review:** Only unconfounded trials demonstrating concealed treatment allocation were included **Assessment of heterogeneity:** Analyses were performed separately according to patient groups and outcome measures. Analyses were also performed separately according to age, gender, diastolic blood pressure and diabetes	**Method used:** Adjusted IC Common control. Results presented in a graph	**Comparisons made:** a. Aspirin+Dip/ctrl (n = 34) b. Aspirin/ctrl (n = 46) c. High-dose aspirin/ctrl (n = 30) d. Medium-dose aspirin/ctrl (n = 19) **Results:** % odds reduction (SD) a. 28% (5) b. 25% (2) c. 21% (4) d. 28% (3) (160–325 mg per day) 26% (11) (<160 mg per day)	**Comparisons made:** e. Aspirin+Dip/aspirin (n = 14) f. High-dose aspirin/medium-dose aspirin (n = 3) **Results:** % odds reduction (SD) e. –1% (9) f. 5% (11)	Indirect (adjusted) evidence used to enhance direct evidence "Such indirect comparisons need to be interpreted more cautiously, for although many of the biases inherent in non-random methods (such as those involving historic controls) are avoided, some potential for bias remains" No statistically significant difference in the results between the indirect and direct comparisons was noted **Appropriate interpretation?** Yes	Additional comparisons were made. Those with greatest sample size presented here
ATC[16] **Validity assessment undertaken within the review:** Only unconfounded trials demonstrating concealed treatment allocation were included **Assessment of heterogeneity:** Separate analyses were undertaken according to patient groups and whether the antiplatelet agent was started before, within 24 hours, or >24 hours after the procedure	**Method used:** Adjusted IC Common control, graphic plot	**Comparisons made:** a. Aspirin+Dip/ctrl (n = 20) b. Aspirin/ctrl (n = 13) **Results:** % odds reduction (SD) a. 40% (6) b. 48% (7)	**Comparisons made:** c. Aspirin+ Dip/aspirin (n = 9) **Results:** % odds reduction (SD) c. –1% (11)	Indirect (adjusted) evidence used to enhance direct evidence "This overview provides some direct and indirect randomised comparisons of the effects of different drug regimens on the prevention of occlusion but finds no evidence of any differences in efficacy. The numbers of patients studied and the numbers of events that occurred were, however, not large enough to exclude some small but real differences in efficacy between different drug regimens" **Appropriate interpretation?** Yes	Additional comparisons were made. Those with greatest sample size presented here

continued

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
ATC[18] **Validity assessment undertaken within the review:** Only unconfounded trials demonstrating concealed treatment allocation were included **Assessment of heterogeneity:** Data from patients undergoing general, traumatic orthopaedic and elective orthopaedic surgery, and high-risk medical patients were analysed separately. Separate analyses were also undertaken for trials in which patients did or did not receive heparin	**Method used:** Adjusted IC using a no-treatment comparison group. Illustrated with plot	**Comparisons made:** On DVT: a. Aspirin+Dip/ctrl (n = 18) b. Aspirin/ctrl (n = 16) On PE: c. Aspirin+Dip/ctrl (n = 18) d. Aspirin/ctrl (n = 21) **Results:** Odds reduction a. 56% b. 23% c. 43% d. 67%	**Comparisons made:** On DVT: e. Aspirin+Dip/aspirin (n = 9) On PE: f. Aspirin+Dip/aspirin (n = 11) **Results:** Odds reduction e. 52%	The discussion of aspirin plus Dip versus aspirin alone was formed on evidence from direct comparisons. The conclusions were cautious Results from the IC support those from direct comparisons **Appropriate interpretation?** Yes	
Lowenthal and Buyse[19] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** χ^2 test for heterogeneity was calculated for each end-point and for each group of trials (aspirin vs placebo, aspirin plus dipyrimadole vs placebo). No significant heterogeneity shown	**Method used:** Adjusted IC, using placebo as comparator. Displayed in graph	**Comparisons made:** a. Aspirin+Dip/placebo (n = 2) b. Aspirin/placebo (n = 9) **Results:** Risk reduction (SD) All deaths: a. 30% (11) b. 10% (8) (ns) Vascular deaths: a. 24% (13) b. –4% (10) (ns) All strokes a. 42% (9) b. 17% (7) (p = 0.007) Fatal strokes: a. 43% (18) b. –10% (21) (p = 0.03) IVE: a. 40% (8) b. 18% (6) (p = 0.007)	**Comparisons made:** c. Aspirin+Dip/aspirin (n = 2) **Results:** Risk reduction (95% CI) All deaths: –19% (–77 to 20%) Vascular deaths: –11% (–78 to 31%) All strokes: 13% (–22 to 38%) Fatal strokes: 2% (–129 to 8%) IVE: 4% (–28 to 28%)	Indirect (adjusted) evidence used to enhance direct evidence. "It must be stressed that these results are based on an indirect comparison between two groups of trials, and may therefore reflect differences in selection criteria or other confounding factors rather than a truly greater treatment effect of combination therapy" "These results *suggest* that the combination therapy of aspirin with dipyridamole may be superior to aspirin alone" **Appropriate interpretation?** Yes	Results subsequently confirmed by larger RCT

continued

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Li Wan Po and Zhang[20] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** Statistical heterogeneity across individual studies tested using the *Q* statistic. If trials were statistically heterogeneous a random effects model was used	**Method used:** Adjusted IC using placebo as comparator	**Comparisons made:** a. Paracetamol + dextroporopxyphene/ placebo (n = 9) b. Paracetamol/placebo (n = 17) **Results:** Mean difference in percentage sum of differences in pain intensity (95% CI) Fixed effect: a. 12.7% (9.2 to 16.2%) b. 9.4% (6.9 to 11.9%) Random effect: a. 13.5% (8.8 to 18.3%) b. 9.4% (6.6 to 12.2%)	**Comparisons made:** c. Paracetamol/placebo (n = 6, only three used in meta-analysis) **Results:** Mean difference in percentage sum of differences in pain intensity (95% CI) c. Fixed effect: 7.3% (–0.2 to 14.9%) Random effects: 7.4% (–0.4 to 15.1%)	Indirect (adjusted) evidence used to enhance direct evidence Authors concluded that "On the basis of data on analgesic efficacy and acute safety in both head to head and indirect comparisons, there is little objective evidence to support prescribing a combination of paracetamol and dextropropoxyphene in preference to paracetamol alone in moderate pain such as that after surgery" **Appropriate interpretation?** Yes	Other outcome measures also presented in the review
Matchar *et al.*[22] **Validity assessment undertaken within the review:** Only RCTs included according to previously published criteria[127] **Assessment of heterogeneity:** No formal test of heterogeneity reported, although it is stated that a random effects method was used as appropriate	**Method used:** Adjusted IC using placebo/ctrl group	**Comparisons made:** Non-valvular atrial fibrillation; a. Warfarin/ctrl (n = 5) b. Aspirin/placebo (n = 2) **Results:** RR for stroke (95% CI) a. 0.33 (0.22 to 0.50) b. 0.67 (0.45 to 0.99)	**Comparisons made:** Non-valvular atrial fibrillation; c. Warfarin/aspirin (n = 1) **Results:** RR for stroke (95% CI) c. 0.34 (0.11 to 0.87)	The results of the IC were used to enhance the findings of the direct comparison **Appropriate interpretation?** Yes	Details on other comparisons and outcome measures are presented in the review

continued

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Moore *et al.*[24] **Validity assessment undertaken within the review:** None reported. Only RCTs with blinded design included **Assessment of heterogeneity:** Tests used not reported. All but two combinations of trials were homogeneous	**Method used:** Adjusted IC using placebo comparator	**Comparisons made:** a. Paracetamol+codeine/ placebo (n = 18) b. Paracetamol/placebo (n = 13) **Results:** >50%maxTOTPAR. Risk ratio (95% CI) a. 2.6 (2.1 to 3.2) b. 1.7 (1.3 to 2.2) Summary statistic (RR): 1.53	**Comparisons made:** c. Paracetamol+ codeine/ paracetamol (n = 11) **Results:** >50%maxTOTPAR. Risk ratio (95% CI) c. 1.19 (0.98 to 1.44)	Results of adjusted IC and direct comparison are discussed separately, with no attempt to combine the results The IC gave a greater estimate than the direct comparison in terms of efficacy of paracetamol plus codeine compared with paracetamol alone **Appropriate interpretation?** Yes	Non-significant results of direct comparison could become significant if the results of the adjusted IC are incorporated
Piccinelli *et al.*[28] **Validity assessment undertaken within the review:** None reported. Only double-blind RCTs of ⩾4 weeks' duration included **Assessment of heterogeneity:** χ^2 test of heterogeneity was undertaken. Fixed effect model used	**Method used:** Adjusted IC	**Comparisons made:** Clomipramine/placebo (n = 9) SSRI/placebo (n = 8) **Results:** Obsessive/compulsive symptoms (effect size = g) (95% CI) a. g = 1.31 (1.15 to 1.47) b. g = 0.47 (0.33 to 0.61)	**Comparisons made:** c. Clomipramine/SSRIs (n = 3) **Results:** c. –0.04 (–0.43 to 0.35)	The authors concluded that "although the increase in improvement rate over placebo was greater for clomipramine than for SSRIs, direct comparison between these drugs showed that they had similar therapeutic efficacy on obsessive–compulsive symptoms" **Appropriate interpretation?** Yes	Considerable difference observed in indirect comparison but not in direct comparison

continued

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Poynard *et al.*[26] **Validity assessment undertaken within the review:** Use of previously validated questionnaire[128] **Assessment of heterogeneity:** χ^2 test of heterogeneity was undertaken. Both fixed and random effect models were used	**Method used:** Adjusted IC	**Comparisons made:** Complete ALT a. 3 MU 12 months/ctrl (n = 7) b. 3 MU 6 months/ctrl (n = 7) Sustained ALT c. 3 MU 12 months/ctrl (n = 5) d. 3 MU 6 months/ctrl (n = 6) **Results:** Response rate (95% CI) a. 48% b. 45% (35 to 55%) Difference in response rate: 3% c. 35% (28 to 43%) d. 21% (13 to 28%) Difference in response rate: 14%	**Comparisons made:** Complete ALT e. 3 MU 12 months/3MU 6 months (n = 4) Sustained ALT f. 3 MU 12 months/ 3 MU 6 months (n = 4) **Results:** e. 11% f. 16% (9 to 23%)	IC used to enhance the results of the direct comparisons **Appropriate interpretation?** Yes	Dose–effect data available in paper

continued

TABLE 19 *Systematic reviews reporting adjusted indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Zhang and Li Wan Po[27] **Validity assessment undertaken within the review:** None reported. Only double-blind RCTs included **Assessment of heterogeneity:** χ^2 test of heterogeneity was undertaken. A random effects model was used when heterogeneity was present	**Method used:** Adjusted IC using placebo comparator	**Comparisons made:** a. Paracetamol+codeine/ placebo (n = 37) b. Paracetamol/placebo (n= 13) c. Paracetamol+caffeine/ placebo (n = 228) d. Paracetamol/placebo (n = 10) **Results:** Difference in TOTPAR% a. d_2 = 23.18 (SE 2.67) b. d_1 = 15.06 (SE 0.90) $d_2 - d_1$ = 8.12 (SE 2.82) c. d_2 = 17.36 (SE 1.89) d. d_1 = 13.91 (SE 1.52) $d_2 - d_1$ = 3.45 (SE 2.43)	**Comparisons made:** e. Paracetamol+ codeine/ paracetamol (n = 13) f. Paracetamol+ caffeine/ paracetamol (n = 10) **Results:** Difference in TOTPAR% e. d = 7.39 (SE 2.43) f. d = 3.97 (SE 1.73)	IC used to support results from direct comparisons "The analgesic efficacy of paracetamol 600 mg was enhanced with the addition of codeine 60 mg (using TOTPAR% as outcome) in both indirect and head-to-head comparisons" **Appropriate interpretation?** Yes	Results for SPID%, ResRR and RemRR also available in the review. The observed difference in TOTPAR was not confirmed using the more clinically meaningful RemRR

ATC, Antiplatelet Trialists' Collaboration; ctrl, control; Dip, dipyridamole; DVT, deep vein thrombosis; IVE, important vascular events; ns, not significant; PE, pulmonary embolism; RemRR, remedication rate ratio; ResRR, response rate ratio; SPID, sum of pain intensity difference.

TABLE 20 *Systematic reviews reporting naive indirect comparisons and direct comparisons*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Marshall and Irvine[21] **Validity assessment undertaken within the review:** 30-point scoring system used.[129] Validity score for each included trial presented. RCTs only included **Assessment of heterogeneity:** Homogeneity within groups of trials confirmed using Breslow–Day test	**Method used:** Naive presentation of pooled response rates across all trials for each treatment	**Comparisons made:** Rectal corticosteroids/placebo or other treatment (n = 16) 5-ASA/other treatment (n = 9) **Results:** Pooled improvement rates by symptomatic, endoscopic and histological criteria: Rectal corticosteroids 77%, 66% and 52%, respectively; 5-ASA 82%, 73% and 66%, respectively	**Comparisons made:** Rectal corticosteroids/5-ASA (n = 7) **Results:** Pooled OR (95% CI) Symptomatic improvement: 1.36 (0.88 to 2.09) Endoscopic improvement: 1.06 (0.61 to 1.85) Histological improvement: 2.27 (1.22 to 4.27)	Emphasis given to results from direct comparison **Appropriate interpretation?** Yes	
Milne *et al.*[23a] **Validity assessment undertaken within the review:** None reported. Comparative clinical studies included **Assessment of heterogeneity:** χ^2 test of heterogeneity was undertaken. No significant heterogeneity was shown	**Method used:** Naive IC Summary statistic of reported adverse events and withdrawals based on data from arms of all trials	**Comparisons made:** a. Roxithromycin/other macrolide or agent commonly used as first line therapy (n = 13) b. Erythromycin/other macrolide or agent commonly used as first line therapy (n = 15) **Results:** Adverse events, rate % (95% CI): a. 10% (8 to 12%) b. 24.8% (22 to 27%) Difference (SE) 14% (1.5%) a. 2.0% (1 to 3%) b. 7.1% (6 to 9%) Difference (SE) 5% (0.045%)	**Comparisons made:** c. Roxithromycin/erythromycin (n = 3) **Results:** No summary statistic for DC alone presented. Adverse events: Roxithromycin 12.8% (10.5 to 17.5%) Erythromycin 27.15% (8 to 51.3%) Withdrawals: Roxithromycin: 1.8% Erythromycin: 2.65%	Authors discussed the possibility of systematic bias due to dissimilar patient groups, but concluded that this is unlikely in this case **Appropriate interpretation?** Unclear	Potential confounding factors discussed and concluded no significant difference between groups in terms of clinical efficacy, age, gender, settings, duration of treatment, indications or year of publication

continued

TABLE 20 *Systematic reviews reporting naive indirect comparisons and direct comparisons (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Pope *et al.*[25] **Validity assessment undertaken within the review:** Studies were assessed for blinding, randomisation and control group. Individual quality assessment for each study not reported **Assessment of heterogeneity:** Heterogeneity was examined and a random effects model used	**Method used:** Naive IC. Adjusted for dietary salt intake	**Comparisons made:** a. Naproxen (n = 4) b. Indomethacin (n = 57) c. Piroxicam (n = 4) d. Sulindac (n = 23) e. Aspirin (n = 4) f. Ibuprofen (n = 6) g. Placebo (n = 10) (n = number of treatment groups for hypertensive patients) **Results:** Mean MAP ±SEM (adjusted for salt intake): b. 3.59 ± 1.12 d. –0.16 ± 1.45 Mean difference in MAP 3.75 ± 1.83 Mean MAP ±SEM (adjusted for salt intake): b. 3.59 ± 1.12 g. –2.59 ± 1.78 Mean difference in MAP 6.18 ± 2.10	**Comparisons made:** h. Indomethacin/sulindac (n=16) i. Indomethacin/placebo (n=11) **Results:** Mean±SEM difference in MAP: a. 4.15±1.00 b. 3.93±1.42	The results of the naive IC were compared with those of direct comparisons when available **Appropriate interpretation?** Yes	Results for other NSAIDs are available in the review

[a] Not identified through DARE.

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Aro[37a] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** No formal test of heterogeneity reported	**Method used:** Adjusted IC Age-standardised	**Comparisons made:** a. PEP+EE/orchidectomy ($n = 1$) b. PEP/orchidectomy ($n = 1$) **Results:** Age-standardised death rate ratio (95% CI) All-cause mortality: a. 2.31 (1.92 to 2.79) b. 1.50 (1.06 to 2.11) CHD deaths: a. 1.51 (1.10 to 2.08) b. 0.17 (0.05 to 0.57)	**Comparisons made:** Not done No trial included in the review directly compared PEP + EE with PEP alone **Results:** Not applicable	Authors concluded that "intramuscular PEP monotherapy is associated with low cardiovascular mortality and with an all-cause and prostatic cancer mortality equal to orchidectomy" **Appropriate interpretation?** No	A subsequent publication[48] highlights the major differences in inclusion criteria between the two studies included in the review. The observed low mortality in PEP alone may be due to flaws in the methodology
Boersma *et al*.[9] **Validity assessment undertaken within the review:** None reported. Only RCTs including ⩾100 patients were included **Assessment of heterogeneity:** Heterogeneity was examined using the Breslow–Day test. Sensitivity analysis was undertaken when heterogeneity existed	**Method used:** Adjusted IC, presented in forest plot Linear and non-linear regression	**Comparisons made:** Time to fibrinolytic therapy (hours:) a. 0–1/ctrl b. ≥ 1–2/ctrl c. ≥ 2–3/ctrl d. ≥ 3–6/ctrl e. ≥ 6–12/ctrl f. ≥ 12–24/ctrl Total $n = 22$ **Results:** Subgroup results. Absolute reduction (SD) in mortality per 1000 a. 65 (14) b. 37 (9) c. 26 (6) d. 29 (5) e. 18 (6) f. 9 (7)	Not done	The authors conclude that "The beneficial effect of fibrinolytic therapy is substantially higher in patients presenting within 2 h after symptom onset compared to those presenting later" The authors do not mention the potential problems associated with ICs. However, in an earlier meta-analysis (upon which this review is based) it is argued that "If patient categories can be arranged in some meaningful order then … may be reasonably reliably informative …" **Appropriate interpretation?** No	Alternative analysis of previous meta-analysis[47]

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Garg and Yusuf[7] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** χ^2 test for heterogeneity was undertaken. No significant heterogeneity shown for total mortality	**Method used:** Adjusted IC	**Comparisons made:** Different ACE inhibitors: a. Benazepril hydrochloride/ctrl (n = 2) b. Captopril/ctrl (n = 6) c. Cilazapril/ctrl (n = 1) d. Enalapril maleate/ctrl (n = 7) e. Lisinopril/ctrl (n = 4) f. Perindopril/ctrl (n = 1) g. Quinapril hydrochloride/ctrl (n = 5) h. Ramipril/ctrl (n = 6) Different duration of follow-up: ≤90 days/ctrl (n = 32) >90 days/ctrl (n = 12) **Results:** Total mortality (OR) Different ACE inhibitors: a. 0.36 b. 0.79 c. 0.12 d. 0.78 e. 0.62 f. 0.14 g. 0.79 h. 0.67 Duration of follow-up: ≤90 days/ctrl 0.56 (0.44, 0.70) > 90 days/ctrl 0.87 (0.75, 1.01)	Not done: no trials making direct comparison between different ACE inhibitors reported in the review	Authors concluded that "similar benefits were observed with several different ACE inhibitors, although data were largely based on enalapril maleate, captopril, ramipril, quinapril hydrochloride and lisinopril" "The greatest effect was seen during the first 3 months, but additional benefit was observed during further treatment" **Appropriate interpretation?** Unclear	It is possible to make within-trial comparison for duration of follow-up and effect, although the conclusion is unlikely to be different

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Holme[35] **Validity assessment undertaken within the review:** None reported. Only RCTs with 6-month follow-up included **Assessment of heterogeneity:** Meta-regression undertaken	**Method used:** Meta-regression (multiple) adjusted by cholesterol changes and baseline risk of CAD	**Comparisons made:** Diet/placebo (n = 12) Hormones/placebo (n = 3) Fibrates/placebo (n = 7) Statins/placebo (n = 3) Other (other drugs/surgery)/placebo (n = 6) **Results:** Total mortality: Diet/fibrates OR 0.975 ($p > 0.05$), $z = 0.84$ Statins/fibrates OR 0.833 ($p < 0.05$), $z = 3.37$	Not done: no trials making direct comparison between different interventions reported in the review	**Appropriate interpretation?** Unclear	
Homocysteine Lowering Trialists' Collaboration[8] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** Heterogeneity across studies assessed by multivariate regression analysis	**Method used:** Adjusted IC using a no-treatment control as common comparator. Illustrated in graph	**Comparisons made:** Different folic acid regimen: a. 1 mg/ctrl (n=5) b. 1–3 mg/ctrl (n=5) c. >3 mg/ctrl (n=4) **Results:** % reduction in blood homocysteine (95% CI): a. 26% (235 to 29%) b. 25% (20 to 29%) c. 255 (21 to 28%)	Not done	Authors suggest a wide range of doses (0.5–5 mg) to be similarly effective **Appropriate interpretation?** Yes	

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Koch *et al.*[34] **Validity assessment undertaken within the review:** Validity assessed using previously published criteria.[130,131] Median quality score for all trials presented. Only RCTs included **Assessment of heterogeneity:** A L'Abbe plot and the *Q* statistic were used to examine heterogeneity	**Method used:** Adjusted IC	**Comparisons made:** H_2 blockers/placebo (nine trials) Misoprostol/placebo (ten trials) **Results:** Rate difference (95% CI) Short-term risk for gastric ulcer: H_2 blockers/placebo: –0.9% (–4.0 to 2.2%) Misoprostol/placebo: –13.3% (–25.7 to –0.9%) Long-term risk for gastric ulcer: H_2 blockers/placebo: –0.3% (–2.9 to 2.2%) Misoprostol/placebo: –8.4% (–17.7 to –1.0%)	Not done: no trials making direct comparison between H_2 blockers and misoprotol reported in the review	Author's claim that "gastric ulcer was found to be significantly reduced by misoprostol, both in short-term and long-term NSAID treatment, but not by H_2 blockers" "We found discrepancies between the results of H_2 blocker and misoprostol trials. Studies on misoprostol admitted subjects with a higher risk for the development of gastric damage" **Appropriate interpretation?** Unclear	Results also presented for gastric lesions, duodenal ulcers and duodenal lesions
Lefering and Neugebauer[39] **Validity assessment undertaken within the review:** A previously published quality checklist was used.[132] A quality score was presented for each trial. Only RCTs included **Assessment of heterogeneity:** A test of heterogeneity was conducted according to Cochran and a random effects model used	**Method used:** Adjusted IC	**Comparisons made:** a. Low-dose corticosteroid/ctrl (n = 5) b. High-dose corticosteroid/ctrl (n = 5) **Results:** Mortality rate % (95% CI) a. –1.9% (–20.0 to 16.2%) b. 3.6% (2.5 to 9.8%)	Not done. No head-to-head trials of high vs low dose available	"Neither the type of steroid used nor the separation into low-dose or high-dose regimen indicated a remarkable difference between the steroid group and control group" **Appropriate interpretation?** Unclear	

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Leizorovicz[30] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** No significant heterogeneity was observed	**Method used:** No formal method used. Comparison of OR, presented on forest plot	**Comparisons made:** a. LMWH hospital/UFH hospital (n = 18) b. LMWH home/UFH hospital (n = 2) **Results:** Recurrent thromboembolic events (OR): a. 0.76 b. 0.79 Mortality: a. 0.66 b. 0.75	Not done. Trials ongoing	"Although this approach reduced the statistical power for each subgroup, efficacy results were similar for recurrence of thromboembolic events or death" **Appropriate interpretation?** Yes	IC also made for route of administration
Moore and McQuay[36] **Validity assessment undertaken within the review:** Validity assessed using the Jadad score[133] All trials scored maximum 5 points **Assessment of heterogeneity:** No formal examination of heterogeneity presented	**Method used:** Adjusted IC, using placebo as comparator. The same placebo group may have been used to compare with different active drugs within trials	**Comparisons made:** a. Codeine (60 mg)/placebo b. Tramadol (50 mg)/placebo c. Tramadol (75 mg)/placebo d. Tramadol (100 mg)/placebo e. Tramadol (150 mg)/placebo f. Acetaminophen (650 mg) plus propoxyphene (100 mg)/placebo g. Aspirin (650 mg) plus codeine (60 mg)/placebo **Results:** RR (95% CI) for patients achieving ⩾50% of %max-TOTPAR Dental pain: c. 1.3 (0.8 to 2.1) d. 2.9 (1.6 to 5.2) e. 2.7 (1.1 to 6.5) f. 3.8 (2.4 to 5.8) g. 4.8 (2.1 to 11.1) h. 4.0 (1.7 to 9.4) i. 3.8 (2.2 to 6.8)	Not done, although possible from data presented	Interpreted as if direct comparisons were made **Appropriate interpretation?** Unclear	Data on postsurgical pain also presented in paper

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Non-small Cell Long Cancer Collaborative Group[32] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** χ^2 test for heterogeneity was undertaken for both between and within chemotherapy categories. When gross heterogeneity was detected, the rationale for combining data was questioned and sources of heterogeneity examined, rather than using a random effects model	**Method used:** Adjusted IC	**Comparisons made:** a. Surgery plus long-term alkylating agent/surgery alone b. Surgery plus cisplatin/surgery alone c. Radical radiotherapy (RT) plus alkylating agents/radical RT alone d. Radical RT plus cisplatin/radical RT alone **Results:** O–E deaths (variance) a. 55.53 (394.74) b. –21.58 (151.83) c. –2.83 (140.23) d. –57.08 (411.18)	Not done: no trials making direct comparison between chemotherapy drugs reported in the review	"Further randomised trials are needed to determine which regimens are the most effective of the modern chemotherapies studied" **Appropriate interpretation?** Yes	Other comparisons presented in article. Only main results extracted
Pignon *et al.*[31] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** Tests of heterogeneity were performed; however, the authors state that "substantial heterogeneity does not invalidate the results of a meta-analysis"	**Method used:** Adjusted IC	**Comparisons made:** a. Early RT/ctrl (n = 7) b. Late RT/ctrl (n = 6) c. With sequential RT/ctrl (n = 8) d. Without sequential RT/ctrl (n = 5) **Results:** OR (95% CI) a. 0.88 (0.78 to 0.98) b. 0.81 (0.69 to 0.94) c. 0.86 (0.75 to 1.00) d. 0.85 (0.75 to 0.96)	Not done. No trials directly comparing early/late, with/without sequential RT available	"Indirect comparison of early with late radiotherapy and of sequential with non-sequential radiotherapy did not reveal any optimal time for treatment" "The selection of an optimal schedule of CT combined with RT that would lead to a major increase in survival with minimal toxicity is the principal challenge raised by our study … .We hope that the results of future trials will settle this question" **Appropriate interpretation?** Yes	

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Poynard *et al.*[38] **Validity assessment undertaken within the review:** The methodological quality of each trial was assessed using a 14-item questionnaire and scored between –2 and 26. A breakdown of the scoring for each trial is presented **Assessment of heterogeneity:** Sensitivity analysis was performed by stratification according to H_2 blocker using both random and fixed effects	**Method used:** Adjusted IC	**Comparisons made:** Lansoprazole/ranitidine or famotidine (n = 5) Other drugs/raniditine or famotidine (n = ?) **Results:** OR (95% CI) for 4-week healing rate in comparison with ranitidine or famotidine Lansoprazole: 2.5 (1.7 to 3.7) Omeprazole: 2.9 (1.5 to 5.7) Ranitidine or famotidine: 0.9 (0.7 to 1.2) Nizatidine: 1.0 (0.8 to 1.4) Cimetidine: 1.8 Sucralfate: 1.0 (0.7 to 1.4)	Lansoprazole directly compared with ranitidine or famotidine. Authors mention an RCT directly comparing lansoprazole versus omeprazole in acute duodenal ulceration; however, this trial is not included in the meta-analysis	The indirect comparison was used to rank efficacy. It was concluded that "there was a significant difference between the efficacy of omeprazole and lansoprazole on the one hand and all the other groups on the other" **Appropriate interpretation?** No	A previous meta-analysis used the same indirect method.[134]
Rossouw[33] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** The authors acknowledge that clinical heterogeneity exists between the included studies. Correlation analyses were performed to assess whether the ORs for disease change were related to baseline or in-trial low-density lipoprotein levels or to relative or absolute differences between treatment and control groups during the trial	**Method used:** Adjusted IC	**Comparisons made:** Various interventions/ctrl: a. Lifestyle/ctrl (n = 3) b. Resins/ctrl (n = 2) c. Statins/ctrl (n = 4) d. Combination/ctrl (n = 5) e. Surgery/ctrl (n = 1) **Results:** OR (95% CI) for cardiovascular events presented in forest plot. Only the details of the overall OR for all interventions presented in the text 0.53 (0.45 to 0.63)	Not done	Authors concluded that "there is no conclusive evidence of a class effect" **Appropriate interpretation?** Yes	

continued

TABLE 21 *Systematic reviews reporting adjusted indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Tramer *et al.*[11] **Validity assessment undertaken within the review:** The Jadad score was used[133] Results not presented **Assessment of heterogeneity:** No formal examination of heterogeneity presented	**Method used:** Adjusted IC	**Comparisons made:** Different doses of droperidol: a. 10 μg kg^{-1} (n = 1) b. 20 μg kg^{-1} (n = 1) c. 50 μg kg^{-1} (n = 2) d. 75 μg kg^{-1} (n = 10) **Results:** Absence of early vomiting: OR (95% CI) a. 1.6 (0.4 to 7.2) b. 1.9 (0.7 to 5) c. 1.5 (0.7 to 3.2) d. 3.3 (2.4 to 4.7)	Not done	The authors' caution about the adjusted IC was mainly due to small sample size. Potential bias was not mentioned "The number of children studied at the lowest doses was small and the confidence intervals wide; nevertheless there would seem to be sufficient information to suggest that the use of submaximal doses is not worthwhile" **Appropriate interpretation?** Unclear	Details of other comparisons and outcome measures presented in the review
Zalcberg *et al.*[29a] **Validity assessment undertaken within the review:** None reported. Only RCTs included **Assessment of heterogeneity:** No formal examination of heterogeneity presented	**Method used:** Adjusted IC Regression analysis was also used	**Comparisons made:** a. ≥ 10 g 5-FU/ctrl (n = 3) b. 8–9 g 5-FU/ctrl (n = 7) c. <8 g 5-FU/ctrl (n = 5) d. Oral CT/ctrl (n = 2) e. 5-FU+levamisole/ctrl (n = 2) f. 5-FU/ctrl (n = 15) **Results:** OR (95% CI) a. 0.71 b. 0.79 c. 0.93 d. 1.04 e. 0.64 (0.49 to 0.85) f. 0.86	Not done. No head-to-head trials presented in the review	"It should be pointed out that the relative effects of 5-FU dose and of levamisole are based on indirect, non-randomised comparisons in this analysis, so that confounding by the type of patients being studied in each trial is a possibility" The authors also conclude that an RCT is required to compare 5-FU+levamisole with 5-FU alone **Appropriate interpretation?** Yes	The dose–response relationship is quite convincing. However, the benefit of additional levamisole to 5-FU is unclear as the trials that included additional levamisole also used higher doses of 5-FU

[a] Not identified through DARE.
CHD, coronary heart disease; CT, chemotherapy; RT, radiotherapy.

TABLE 22 *Systematic reviews reporting naive indirect comparisons only*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Bansal and Beto[42] **Validity assessment undertaken within the review:** None reported. Prospective controlled trials with treatment allocation by random assignment or consecutive enrolment **Assessment of heterogeneity:** Heterogeneity was assessed and a random effects model used. The appropriateness of pooling was also assessed by comparing the results between those trials with matched control and all trials	**Method used:** Naive IC of results of different arms across studies The appropriateness of pooling was examined by two methods (*Z*-score and heterogeneity test), but the results of the tests were not presented	**Comparisons made:** Treatment arms across studies were pooled to compare various immunosuppressive agents plus oral prednisone with prednisone alone (n = ?) **Results:** Absolute risk difference (95% CI) for all immunosuppressive agents with prednisone/prednisone alone Total mortality: 13.2% (2.5 to 23.9%) ESRD: 12.9% (2.2 to 23.6%)	Not done	The conclusion was based on a naive IC without any effort to adjust for potential bias and confounding. Results were interpreted as though DC undertaken with no discussion of potential biases **Appropriate interpretation?** No	The review included RCTs or quasi-RCTs, but the results of the studies were used to make between-study comparisons, losing the power/rigour of randomisation
Chiba *et al.*[40] **Validity assessment undertaken within the review:** Blinding and method of randomisation were the main quality items assessed. Only single- or double-blind RCTs were included **Assessment of heterogeneity:** No examination of heterogeneity reported	**Method used:** Naive IC For each drug class linear regression analysis estimated the average percentage of patients who were healed and heartburn free per week	**Comparisons made:** 43 studies used to compare PPIs (omeprazole, lansoprazole and pantoprazole), H_2RAs (cimetidine, nizatidine, rantidine and famotidine), sucralfate, prokinetics, placebo and other **Results:** Mean overall healing proportion: PPIs: 83.6 ± 11.4% H_2RA: 51 ± 17.1% Sucralfate: 39.2 ± 22.4% Placebo: 28.2 ± 15.6%	Not done but possible	The authors conclude that "more complete esophagitis healing and heartburn relief is observed with PPIs vs H_2-RAs" **Appropriate interpretation?** No	The review includes only single- or double-blind RCTs, but no within-study comparison was made It is not clear whether it is a standard error or standard deviation following the estimated overall proportion

continued

TABLE 22 *Systematic reviews reporting naive indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Conlin *et al.*[46] **Validity assessment undertaken within the review:** None reported. Only double-blind RCTs were included **Assessment of heterogeneity:** No examination of heterogeneity reported	**Method used:** Naive IC Analysis based on treatment arms. The absolute (non-placebo corrected) weighted average blood pressure reduction was calculated for each AIIA	**Comparisons made:** 43 trials were used to compare the efficacy of losartan, valsartan, irbesartan and candesartan **Results:** The absolute weighted average reductions in diastolic and systolic blood pressure were comparable for all AIIAs	Not done, although possible to do so	Authors conclude that "this analysis suggests that AIIA lower blood pressure with similar efficacy when administered at their usual doses for the treatment of hypertension" **Appropriate interpretation?** No	The review includes only RCTs, but no within-study comparison was made. Overestimation of blood pressure reduction because of regression to mean effect, owing to the ignoring of the placebo groups
Coulter *et al.*[41] **Validity assessment undertaken within the review:** None reported. Only RCTs were included **Assessment of heterogeneity:** Clinical heterogeneity between trials was noted. No tests of heterogeneity were reported	**Method used:** Naive IC	**Comparisons made:** 31 studies used to compare the efficacy of a variety of drugs used to treat menorrhagia **Results:** Drugs listed according to percentage reduction in menstrual blood loss	Not done, although possible to do so	Authors mention limitations such as lack of placebo controls, but claim the review of RCTs is satisfactory for comparing relative efficacy of drugs **Appropriate interpretation?** Unclear	The review includes only RCTs, but no within-study comparison was made. Half of the trials included a placebo group, but results of the placebo controls were not reported

continued

TABLE 22 *Systematic reviews reporting naive indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Felson *et al.*[6] **Validity assessment undertaken within the review:** Modification of a previously published checklist.[135] Only RCTs were included **Assessment of heterogeneity:** χ^2 test of heterogeneity was undertaken and a random effects model used when appropriate	**Method used:** Naive IC Adjustment was made for some covariates, but not clear how adjustment was made	**Comparisons made:** Trials of second line drugs to treat rheumatoid arthritis were included: a. Placebo (n = 22) b. Antimalarial drugs (n = 11) c. Auranofin (n = 23) d. Injectable gold (n = 29) e. Methotrexate (n = 7) f. DP (n = 19) g. SSZ (n = 6) **Results:** Efficacy: Auranofin was found to be significantly weaker than methotrexate, injectable gold, DP and SSZ, and slightly, but not significantly weaker than antimalarial agents Toxicity: Injectable gold had higher toxicity rates and higher total dropout than any other drug	**Comparisons made:** Not done, although possible to do so	Within-study comparisons ignored Information from all outcome measures and from each trial combined into a composite measure of treatment effect **Appropriate interpretation?** No	This review was been quoted by Imperiale and Speroff as proof of using naive ICs of RCT data Direct comparison studies available

continued

TABLE 22 *Systematic reviews reporting naive indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Imperiale and Speroff[45] **Validity assessment undertaken within the review:** Only RCTs included. Trials assessed on inclusion/exclusion criteria; baseline similarity; blind administration of intervention; description of co-interventions; drop-outs and withdrawals. Summary score (max. 10) assigned to each trial **Assessment of heterogeneity:** Analysis based on premise that treatment groups were clinically homogenous. Random effects model used	**Method used:** Naive IC based on premise that the treatment groups were clinically homogeneous in composition	**Comparisons made:** a. LMWH/ctrl (n = 20) b. Warfarin/ctrl (n = 10) c. Compression stockings/ctrl (n = 6) **Results:** NNT (95% CI) for all DVT a. 3.2 (2.9 to 3.7) b. 4.3 (3.0 to 8.0) c. 3.9 (3.1 to 5.1)	Not done	**Appropriate interpretation?** No	Authors give the impression that their conclusions were based on evidence from RCTs, even though only between-study comparisons were made Quotes review by Felson *et al.*[6] to support use of IC method More outcomes and comparisons available in the review
Schmieder *et al.*[43] **Validity assessment undertaken within the review:** None reported. Only double-blind RCTs included **Assessment of heterogeneity:** Homogeneity of data could not be confirmed for all treatment arms, so sensitivity analysis was undertaken	**Method used:** Naive IC All treatment arms of the same drug were combined	**Comparisons made:** a. Diuretics (n = 13) b. β-Blockers (n = 21) c. Calcium-channel blockers (n = 19) d. ACE inhibitors (n = 18) e. Placebo (n = 13) **Results:** Mean (SD) % decrease in systolic/diastolic blood pressure a. 10.7 (1.8)/13.1 (3.5) b. 12.8 (3.8)/15.4 (2.8) c. 10.3 (3.6)/13.4 (2.3) d. 11.9 (4.6)/13.2 (5.4)	Not done, although possible to do so	The authors discuss the importance of randomisation and scientific quality of studies. However, they ignore within-study comparisons and make no attempt to adjust the IC according to a common placebo group **Appropriate interpretation?** No	See updated review by the same authors[136]

continued

TABLE 22 *Systematic reviews reporting naive indirect comparisons only (cont'd)*

Study	Method of IC	Indirect comparison	Direct comparison	Description of interpretation	Comments
Unge and Berstad[44] **Validity assessment undertaken within the review:** None reported. Study designs unclear **Assessment of heterogeneity:** Sensitivity analysis on major differences in dose, dosage and duration was performed	**Method used:** Naive IC	**Comparisons made:** 17 different treatment groups For example: a. Omeprazole + amoxycillin + clarithromycin (n = 59) b. Bismuth + nitroimidazole + tetracycline (n = 87) **Results:** Eradication rate: a. 87% (range 72–100% b. 82% (range 43–100%)	Not done, although possible to do so	The authors conclude that "omeprazole/clarithromycin based triple regimens are the most effective anti-*H. pylori* therapeutic strategy, slightly superior to bismuth triple regimens" **Appropriate interpretation?** No	Observational studies and RCTs are included in the review. Details of other comparisons are presented in the review

DP, D-penicillamine; SSZ, sulfasalazine.

Appendix 4

List of excluded reviews

The following reviews were identified from the searches as potentially including indirect comparisons or both direct and indirect comparison of competing interventions. However, after assessment they did not include suitable data and so were excluded.

Arriagada R, Pignon JP, Ihde DC, Johnson DH, Perry MC, Souhami RL, *et al*. Effect of thoracic radiotherapy on mortality in limited small cell lung cancer. A meta-analysis of 13 randomized trials among 2,140 patients. *Anticancer Res* 1994; **14**(1B):333–5.

van Balkom AJ, Nauta MC, Bakker A. Meta-analysis on the treatment of panic disorder with agoraphobia; review and re-examination. *Clin Psychol Psychother* 1995;**2**:1–14.

Bhansali M, Vaidya J, Bhatt R, Patil P, Badwe R, Desai P. Chemotherapy for carcinoma of the esophagus: a comparison of evidence from meta-analyses of randomized trials and of historical control studies. *Ann Oncol* 1996;**7**:355–9.

Boyer W. Serotonin uptake inhibitors are superior to imipramine and alprazolam in alleviating panic attacks: a meta-analysis. *Int Clin Psychopharmacol* 1995;**10**:45–9.

Chang D, Wilson S. Meta-analysis of the clinical outcome of carbapenem monotherapy in the adjunctive treatment of intra-abdominal infections. *Am J Surg* 1997;**174**:284–90.

Childhood ALL Collaborative Group. Duration and intensity of maintenance chemotherapy in acute lymphoblastic leukaemia: overview of 42 trials involving 12 000 randomised children. *Lancet* 1996;**347**:1783–8.

de Craen A, Di Giulio G, Lampe-Schoenmaechers A, Kessels A, Kleijnen J. Analgesic efficacy and safety of paracetamol-codeine combinations versus paracetamol alone: a systematic review. *BMJ* 1996;**313**:321–5.

Droitcour J, Silberman G, Chelimsky E. A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *Int J Technol Assess Health Care* 1993; **9**:440–9.

Eriksson S, Langstrom G, Rikner L, Carlsson R, Naesdal J. Omeprazole and H_2 receptor antagonists in the acute treatment of duodenal ulcer, gastric ulcer and reflux oesophagitis: a meta-analysis. *Eur J Gastroenterol Hepatol* 1995; **7**:467–75.

Golzari H, Cebul R, Bahler R. Atrial fibrillation: restoration and maintenance of sinus rhythm and indications for anticoagulation therapy. *Ann Intern Med* 1996;**125**:311–23.

Halliday H. Overview of clinical trials comparing natural and synthetic surfactants. *Biol Neonate* 1995;**67**(Suppl):32–47.

Held P, Yusuf S. Calcium anatagonists in the treatment of ischemic heart disease: myocardial infarction. *Coron Artery Dis* 1994;**5**:21–6.

Hoes A, Grobbee D, Lubsen J. Does drug treatment improve survival? Reconciling the trials in mild-to-moderate hypertension. *J Hypertens* 1995;**13**:805–11.

Hoes A, Grobbee D, Peet T, Lubsen J. Do non-potassium-sparing diuretics increase the risk of sudden cardiac death in hypertensive patients? *Drugs* 1994;**47**:711–33.

Hooks M. Tacrolimus, a new immunosuppressant – a review of the literature. *Ann Pharmacother* 1994;**28**:501–11.

Koes B, Assendelft W, van der Heijden G, Bouter L. Spinal manipulation for low back pain. An updated systematic review of randomized clinical trials. *Spine* 1996;**21**:2860–71.

Linde K, Ramirez G, Mulrow C, Pauls A, Weidenhammer W, Melchart D. St John's wort for depression – an overview and meta-analysis of randomised clinical trials. *BMJ* 1996;**313**:253–8.

MacRae H, McLeod R. Comparison of hemorroidal treatment modalities: a meta-analysis. *Dis Colon Rectum* 1995;**38**:687–94.

McQuay H, Carroll D, Jadad AR, Wiffen P, Moore A. Anticonvulsant drugs for management of pain: a systematic review. *BMJ* 1995; **311**:1047–52.

McQuay H, Tramer M, Nye B, Carroll D, Wiffen P, Moore R. A systematic review of antidepressants in neuropathic pain. *Pain* 1996;**68**:217–27.

Meunier F, Paesmans M, Autier P. Value of antifungal prophylaxis with antifungal drugs against oropharyngeal candidiasis in cancer patients. *Eur J Cancer* 1994;**30B**:196–9.

Ofman J, Koretz R. Clinical economics review: nutritional support. *Aliment Pharmacol Ther* 1997; **11**:453–71.

Patrono C, Roth GJ. Aspirin in ischemic cerebrovascular disease. How strong is the case for a different dosing regimen? *Stroke* 1996;**27**:756–60.

Piccinelli M, Pini S, Bellantuono C, Wilkinson G. Efficacy of drug treatment in obsessive–compulsive disorder: a meta-analytic review. *Br J Psychiatry* 1995;**166**:424–43.

Rahlfs V, Macciocchi A, Monti T. Brodimoprim in upper respiratory tract infections. Two meta-analyses of randomised, controlled clinical trials in acute sinusitis and otitis media. *Clinical Drug Investigation* 1996;**11**:65–76.

Rains C, Noble S, Faulds D. Sulfasalazine. A review of its pharmacological properties and therapeutic efficacy in the treatment of rheumatoid arthritis. *Drugs* 1995;**50**:137–56.

Riedemann P, Bersinic S, Cuddy L, Torrance G, Tugwell P. A study to determine the efficacy and safety of tenoxicam versus piroxicam, diclofenac and indomethacin in patients with osteoarthritis: a meta-analysis. *J Rheumatol* 1993;**20**:2095–103.

Tramonte S, Brand M, Mulrow C, Amato M, O'Keefe M, Ramirez G. The treatment of chronic constipation in adults. A systematic review. *J Gen Intern Med* 1997;**12**:15–24.

Vaitkus PT, Berlin JA, Schwartz JS, Barnathan ES. Stroke complicating acute myocardial infarction. A meta-analysis of risk modification by anticoagulation and thrombolytic therapy. *Arch Intern Med* 1992;**152**:2020–4.

Voogel A, van der Meulen J, van Montfrans G. Effects of antihypertensive drugs on the circadian blood pressure profile. *J Cardiovasc Pharmacol* 1996;**28**:463–9.

Wade C, Kramer G, Grady J, Fabian T, Younes R. Efficacy of hypertonic 7.5% saline and 6% dextran-70 in treating trauma: a meta-analysis of controlled clinical studies. *Surgery* 1997; **122**:609–16.

Wood M. The comparative efficacy and safety of teicoplanin and vancomycin. *J Antimicrob Chemother* 1996;**37**:209–22.

Appendix 5

Search strategy development

Search strategy 1

The process involved an initial 'One Search' on Dialog to gauge the amount of literature on indirect comparisons and to identify suitable databases for future searching (search strategy 1)

Search strategy 1:

S1 randomized controlled trials/ab,ti,de
S2 trial?/ab,ti,de
S3 s1 or s2
S4 (indirect(2w)comparison?)/ab,ti,de
S5 (direct(2w)comparison?)/ab,ti,de
S6 (indirect(2w)evaluat?)/ab,ti,de
S7 (direct(2w)evaluat?)/ab,ti,de
S8 (treatment(2w)arm?)/ab,ti,de
S9 (compet?(2w)technolog?)/ab,ti,de
S10 (compet?(2w)intervention?)/ab,ti,de
S11 s4:s10
S12 s3 and s11

Search strategy 1 was run across all the databases listed below up to October 1997 and retrieved records as follows:

Database	**Records retrieved**
MEDLINE	723
ERIC	1
PsycINFO	29
EMBASE	901
Dissertation Abstracts	33
MathSci	8

Search strategy 2

The next step was to run a strategy on MEDLINE via Ovid (Search strategy 2) to retrieve the full text of each record to help in the process of identifying suitable terms for inclusion in the final search strategy. This strategy identified 759 records on MEDLINE when run for the period 1966 to October 1997.

Search strategy 2:

1 exp RANDOMIZED CONTROLLED TRIALS/
2 trial$.tw.
3 1 or 2
4 (indirect adj2 comparison$).tw.
5 (direct adj2 comparison$).tw.
6 (indirect adj2 evaluat$).tw.
7 (direct adj2 evaluat$).tw.
8 (treatment adj2 arm$).tw.
9 (compet$ adj2 technolog$).tw.
10 (compet$ adj2 intervention$).tw.
11 4 or 5 or 6 or 7 or 8 or 9 or 10
12 3 and 11

Search strategy 3

After further consideration of how 'competing interventions' or indirect comparisons were described in individual studies and which MeSH headings had been used to index records, the strategy was then further developed (search strategy 3).

Search strategy 3 was run for the year 1997 and retrieved 1690 records. When these records were examined by the reviewers it was found that many of them were reports of single RCTs, rather than discussion of the methodology of RCTs.

Search strategy 3:

1 RANDOMIZED CONTROLLED TRIALS/
2 controlled clinical trials.sh.
3 CLINICAL TRIALS/
4 clinical trials.tw.
5 trial$.tw.
6 meta-analysis.sh.
7 meta-analysis.tw.
8 metaanalys$.tw.
9 (meta adj analys$).tw.
10 RESEARCH DESIGN/
11 data interpretation,statistical.sh.
12 models,statistical.sh.
13 (indirect adj2 comparison$).tw.
14 (direct adj2 comparison$).tw.
15 (indirect adj2 evaluat$).tw.
16 (direct adj2 evaluat$).tw.
17 (compet$ adj2 technolog$).tw.
18 (compet$ adj2 intervention$).tw.
19 (treatment adj2 arm$).tw.
20 (treatment adj2 group$).tw.
21 (randomi$ adj2 group$).tw.
22 (randomi$ adj2 comparison$).tw.

23 (therapeutic adj2 arm).tw.
24 (therapeutic adj2 arms).tw.
25 (study adj2 arm).tw.
26 (study adj2 arms).tw.
27 4 limb study.tw.
28 four limb study.tw.
29 (trial adj2 arm).tw.
30 (trial adj2 design).tw.
31 (placebo adj2 arm).tw.
32 (preventive adj2 arm).tw.
33 (preventative adj2 arm).tw.
34 or/1-9
35 or/10-33
36 34 and 35

Search strategy 4

In an attempt to remove records describing single studies from the search results the strategy was amended. The desired outcome was to retrieve records in which there was reference to trials and meta-analysis and one of the possible free text terms used for describing competing interventions/indirect comparisons (search strategy 4).

Search strategy 4:

1 RANDOMIZED CONTROLLED TRIALS/
2 controlled clinical trials.sh.
3 CLINICAL TRIALS/
4 clinical trials.tw.
5 trial$.tw.
6 meta-analysis.sh.
7 meta-analysis.tw.
8 metaanalys$.tw.
9 (meta adj analys$).tw.
10 RESEARCH DESIGN/
11 data interpretation,statistical.sh.
12 models,statistical.sh.
13 (indirect adj2 comparison$).tw.
14 (direct adj2 comparison$).tw.
15 (indirect adj2 evaluat$).tw.
16 (direct adj2 evaluat$).tw.
17 (compet$ adj2 technolog$).tw.
18 (compet$ adj2 intervention$).tw.
19 (treatment adj2 arm$).tw.
20 (treatment adj2 group$).tw.
21 (randomi$ adj2 group$).tw.
22 (randomi$ adj2 comparison$).tw.
23 (therapeutic adj2 arm).tw.
24 (therapeutic adj2 arms).tw.
25 (study adj2 arm).tw.
26 (study adj2 arms).tw.
27 4 limb study.tw.
28 four limb study.tw.
29 (trial adj2 arm).tw.
30 (trial adj2 design).tw.
31 (placebo adj2 arm).tw.
32 (preventive adj2 arm).tw.
33 (preventative adj2 arm).tw.
34 or/1-5
35 or/6-9
36 or/10-33
37 34 and 35 and 36

Search strategy 5

Search strategy 4 was run on MEDLINE for the period 1995 to December 1998 and retrieved 238 records. There was still some uncertainty, however, about the recall of the search strategy, and in an attempt to improve this additional free text terms were included in the third facet of the strategy (search strategy 5). The inclusion of the additional terms resulted in the retrieval of some additional 131 records.

Search strategy 5:

1 RANDOMIZED CONTROLLED TRIALS/
2 controlled clinical trials
3 CLINICAL TRIALS/
4 clinical trials.tw.
5 trial$.tw.
6 1 or 2 or 3 or 4 or 5
7 meta-analysis.sh.
8 meta-analysis.tw.
9 metaanalys$.tw.
10 (meta adj analys$).tw.
11 7 or 8 or 9 or 10
12 RESEARCH DESIGN/
13 data interpretation, statistical.sh.
14 models, statistical.sh.
15 (indirect adj2 comparison$).tw.
16 (direct adj2 comparison$).tw.
17 (indirect adj2 evaluat$).tw.
18 (direct adj2 evaluat$).tw.
19 (compet$ adj2 technolog$).tw.
20 (compet$ adj2intervention$).tw.
21 (treatment adj2 arm$).tw.
22 (treatment adj2group$).tw.
23 (randomi$ adj2group$).tw.
24 (randomi$ adj2comparison$).tw.
25 (therapeutic adj2 arm).tw.
26 (therapeutic adj2 arms).tw.
27 (study adj2 arm).tw.
28 (study adj2 arms).tw.
29 4 limb study.tw.
30 four limb study
31 (trial adj2 arm).tw.
32 (trial adj2 design).tw.

33 (placebo adj2 arm).tw.
34 (preventive adj2arm).tw.
35 (preventative adj2 arm).tw.
36 multiple arm study.tw.
37 multiple arms study.tw.
38 multiple arm studies.tw.
39 multiple arms studies.tw.
40 multiple arm.tw.
41 multiple arms.tw.
42 multi arm.tw.
43 multi arms.tw.
44 (multi adj2 arm).tw.
45 (multi adj2 arms).tw.
46 (multiple adj2 arm).tw.
47 (multiple adj2 arms).tw.
48 ((three arm or three arms or 3 arm or 3 arms or three limb or three limbs or 3 limb or 3limbs) adj5 (trial$ or stud$ or random$))
49 ((four arm or four arms or 4 arm or 4 arms or four limb or four limbs or 4 limb or 4 limbs) adj5 (trial$ or stud$ or random$))
50 (competing adj2 therap$).tw.
51 (multi$ adj2 (study or studies)).tw.
52 or/12-51
53 6 and 11 and 52

Before 1989, records in MEDLINE referring to meta-analysis were indexed using the terms: Outcome and Process Assessment (1977–1979), Follow-Up Studies (1977–1979), Research (1980–1982), Research Design (1980–1988) and Statistics (1980–1988). It was not clear what effect this would have on running the existing strategy on these earlier years, so various trial strategies were undertaken. These included (1) omitting the meta-analysis facet from the strategy entirely; (2) replacing the meta-analysis terms with the terms research, research design, follow-up studies, statistics, outcome and process assessment using the explosion facility; and (3) replacing the meta-analysis terms with the terms research, research design, follow-up studies, statistics, outcome and process assessment not using the explosion facility. None of these search strategies resulted in any additional studies being identified. Strategy 5 was therefore used as a basis from which to develop strategies to use in other databases with amendments as appropriate regarding thesaurus terms and subject indexing. The MEDLINE search was run from 1966 to March 1999. It was updated to include records published by February 2001.

Appendix 6

Additional electronic search strategies

PsycLIT

The PsycLIT database was searched via SilverPlatter (1887 to March 1999) and 287 records were retrieved. The search was updated to include records published up to February 2001, identifying a further 40 records. The search strategy used was:

1 randomized controlled trials
2 randomi* control* trial*
3 control* clinical trial*
4 clinical trial*
5 trial*
6 exact{EMPIRICAL-STUDY} in PT
7 1 or 2 or 3 or 4 or 5 or 6
8 multiple arm study
9 multiple arms study
10 multiple arms studies
11 multiple arms studies
12 multiple arm
13 multiple arms
14 multi arm
15 multi arms
16 multi near2 arm
17 multi near2 arms
18 multiple near2 arm
19 multiple near2 arms
20 three arm or three arms or 3 arm or 3 arms or three limb or three limbs or 3 limb or 3 limbs
21 four arm or four arms or 4 arm or 4 arms or four limb or four limbs or 4 limb or 4 limbs
22 multi* near3 (study or studies)
23 8 or 9 or 10 or 11or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24 7 or 23
25 explode META-ANALYSIS
26 meta-analy*
27 metaanaly*
28 meta near analy*
29 25 or 26 or 27 or 28
30 24 and 29
31 explode EXPERIMENTAL-DESIGN
32 explode STATISTICAL-ANALYSIS
33 indirect near2 comparison*
34 direct near2 comparison*
35 indirect near2 evaluat*
36 direct near2 evaluat*
37 compet* near2 technolog*
38 compet* near2 intervention*
39 treatment near2 arm*
40 treatment near2 group*
41 randomi* near2 group*
42 randomi* near2 comparison*
43 therapeutic near2 arm*
44 study near2 arm*
45 trial near2 arm*
46 trial near2 design*
47 placebo near2 arm*
48 preventive near2 arm*
49 preventative near2 arm*
50 competing near2 therap*
51 indirect near evaluat*
52 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51
53 30 and 52

ERIC

The ERIC database was searched using the Ovid interface via BIDS (1966 to February 1999) and 72 records were retrieved. The search strategy used was:

1 randomized controlled trial.tw.
2 randomi#ed control? trial?.tw.
3 control? clinical trial?.tw.
4 clinical trial?.tw.
5 trial?.tw.
6 exp MATCHED GROUPS/
7 exp EXPERIMENTAL GROUPS/
8 1 or 2 or 4 or 5 or 6 or 7
9 multiple arm study.tw.
10 multiple arms study.tw.
11 multiple arm studies.tw.
12 multiple arms studies.tw.
13 multiple arm.tw.
14 multiple arms.tw.
15 multi arm.tw.
16 multi arms.tw.
17 (multi adj2 arm).tw.
18 (multi adj2 arms).tw.
19 (multiple adj2 arm).tw.
20 (multiple adj2 arms).tw.
21 (three arm or three rms or 3 arm or 3 arms or three limb or three limbs or 3 limb).mp. or 3 limbs.tw. [mp=abstract, title, heading word, identifiers]

22 (three arm or three arms or 3 arm or 3 arms or three limb or three limbs or 3 limb).mp. or 3 limbs.tw. [mp=abstract, title, heading word, identifiers]
23 (four arm or four arms or 4 arm or 4 arms or four limb or four limbs or 4 limb).mp. or 4 limbs.tw. [mp=abstract, title, heading word, identifiers]
24 (multi? adj3 (study or studies)).tw.
25 8 or 23 or 24
26 meta-analysis.tw.
27 (meta adj analysis).tw.
28 metaanalysis.tw.
29 meta-analytic.tw.
30 (meta adj analytic).tw.
31 metaanalytic.tw.
32 exp META ANALYSIS/
33 exp LITERATURE REVIEWS
34 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
35 25 and 34

EMBASE

The EMBASE database was searched using the Silverplatter interface (1980 to April 1999) and 282 records were retrieved. The search strategy used was:

1 explode RANDOMIZED-CONTROLLED-TRIAL/ all subheadings
2 explode CONTROLLED-STUDY/ all subheadings
3 explode CLINICAL-TRIAL/ all subheadings
4 trial*
5 clinical trials
6 1 or 2 or 3 or 4 or 5
7 multiple arm study
8 multiple arms study
9 multiple arm studies
10 multiple arms studies
11 multiple arm
12 multiple arms
13 multi arm
14 multi arms
15 multi near2 arm
16 multi near2 arms
17 multiple near2 arm
18 multiple near2 arms
19 three arm or three arms or 3 arm or 3 arms or three limb or three limbs or 3 limb or 3 limbs
20 #19 near5 (trial* or stud* or random*)
21 four arm or four arms or 4 arm or 4 arms or four limb or four limbs or 4 limb or 4 limbs
22 #21 near5 (trial* or stud* or random*)
23 multi* near3 (study or studies)
24 #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #20 or #22 or #23
25 #6 or #24
26 explode META-ANALYSIS/ all subheadings
27 meta-analysis
28 metaanalys*
29 meta near2 analys*
30 #26 or #27 or #28 or #29
31 #25 and #30
32 indirect near2 comparison*
33 direct near2 comparison*
34 indirect near2 evaluat*
35 direct near2 evaluat*
36 compet* near2 technolog*
37 compet* near2 intervention*
38 treatment near2 arm*
39 treatment near2 group*
40 randomi* near2 group*
41 randomi* near2 comparison*
42 therapeutic near2 arm
43 therapeutic near2 arms
44 study near2 arm
45 study near2 arms
46 trial near2 arm
47 trial near2 design
48 placebo near2 arm
49 preventive near2 arm
50 preventative near2 arm
51 competing near2 therap*
52 #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51
53 #31 and #52

MathSci

The MathSci database was searched using Dialog (up to September 1999) and 1817 records were retrieved. The search strategy used was:

1 s research(W)design
2 s statist?(W)models
3 s indirect(2W)comparison?
4 s direct(2W)comparison?
5 s indirect(2W)evaluat?
6 s direct(2W)evaluat?
7 s compet?(2W)technolog?
8 s compet?(2W)intervention?
9 s treatment(2W)arm?
10 s treatment(2W)group?
11 s randomi?(2W)group?
12 s randomi?(2W)comparison?
13 s therapeutic(2W)arm
14 s therapeutic(2W)arms

15 s study(2W)arm
16 s study(2W)arms
17 s 4(W)limb(W)study
18 s four(W)limb(W)study
19 s trial(2W)arm
20 s trial(2W)design
21 s placebo(2W)arm
22 s preventive(2W)arm
23 s preventative(2W)arm
24 s multiple(W)arm(W)study
25 s multiple(W)arms(W)study
26 s multiple(W)arm(W)studies
27 s multiple(W)arms(W)studies
28 s multiple(W)arm
29 s multiple(W)arms
30 s multi(W)arm
31 s multi(W)arms
32 s multi(2W)arm
33 s multi(2W)arms
34 s multiple(2W)arm
35 s multiple(2W)arms
36 s ((three(W)arm) or (three(W)arms) or (3(W)arm) or (3(W)arms) or (three(W)limb) or (three(W)limbs) or (3(W) limb) or (3(W)limbs)) (5W) (trial? or stud? or random?)
37 s ((four(W)arm) or (four(W)arms) or (4(W)arm) or (4(W)arms) or (four(W)limb) or (four(W)limbs) or (4(W)limb) or (4(W)limbs)) (5W) (trial? or stud? or random?)
38 s competing(2W)therap?
39 s multitreatment(2W)(study or studies)
40 s s1:s39

Appendix 7

Illustrative analyses

Data set

Several of the possible analyses are illustrated using a set of trials of antithrombotic therapy to prevent strokes in patients with atrial fibrillation[150] (*Table 23*). The emphasis is on methods that can be applied using widely available software. As these trials are used for illustration only the impact of certain criticisms of the systematic review has been disregarded.[151]

From this set of data subsets of trials are taken to illustrate:

- an indirect comparison of BvC using A as a common control (studies 3, 4 and 5 are two-armed comparisons of AvB, studies; 7, 8, 9 and 10 are two-armed comparisons of AvC)
- a combination of the above indirect comparison with direct two-armed trials to estimate BvC (trial 11 is a two-armed trial of BvC, which is combined with the indirect comparison using the seven trials listed above)
- a combination of the above indirect and direct comparisons with multiarmed trials of A, B and C (trials 1, 2 and 6 combined with the eight trials listed above)
- a combination of all 15 trials and all treatments.

These subsets of trials are created for the purpose of illustrating sequentially more complex alternative models, and the results should not be regarded as definitive analyses of the above data set.

Analyses

Analyses were undertaken using Stata version 8 for adjusted indirect comparisons and logistic regression, and using the PROC NLMIXED procedure in SAS version 9.1 for the mixed models. The SAS procedure is chosen for presentation in preference over the Stata gllamm command (which can also be used for fitting mixed models), for reasons of computational speed.

TABLE 23 *Event rates for stroke in trials of antithrombotic therapy for patients with atrial fibrillation (from Hart* et al.*[150])*

	Trial	Placebo	Adjusted-dose warfarin	Aspirin	Low- or fixed-dose warfarin	Low- or fixed-dose warfarin + aspirin
		A	B	C	D	E
1	AFASAK	19/336	9/335	16/336		
2	SPAF	19/211	8/210	25/552		
3	BAATAF	13/208	3/212			
4	CAFA	9/191	6/187			
5	SPINAF	23/290	7/281			
6	EAFT i[a]	50/214	20/225	49/230[a]		
7	EAFT ii[a]	40/164		39/174[a]		
8	ESPS II	23/107		17/104		
9	LASAF	6/182		5/194[b]		
10	UK-TIA	8/30		8/34[b]		
11	SPAF II		19/358	21/357		
12	AFASAK II		11/170	9/169	14/167	11/171
13	PATAF		3/131	4/141	4/122	
14	SPAF III		14/523			48/521
15	MWNAF	1/153			5/150	

[a] EAFT[152] included two clinical subgroups randomised to different treatment options according to eligibility for anticoagulation. The results for aspirin were combined in the publication (88/404); here, the data have been split by making the event rates equal.
[b] Combination of two groups randomised to different doses of aspirin.

TABLE 24 *Example data set for indirect comparisons analysis*

ID	Trial	event_c	sample_c	event_t	sample_t	compb
1	BAATAF	3	212	13	208	1
2	CAFA	6	187	9	191	1
3	SPINAF	7	281	23	290	1
4	EAFT ii	39	174	40	164	0
5	ESPS II	17	104	23	107	0
6	LASAF	5	194	6	182	0
7	UK-TIA	8	34	8	30	0

Format of data sets

Four different formats are required for alternative analyses:

(a) One row per trial (see *Table 24*)

Each row states for treatment and control groups the numbers of events (event_t, event_c) and the number of participants (sample_t, sample_c).

This data structure is only suitable for two-armed trials. For adjusted indirect comparisons (AICs) meta-analyses are undertaken on different subsets of the data for each component meta-analysis indicated by an additional variable [here a variable compb is used, indicating whether A was compared with B (value 1) or C (value 0)]. Here, the 'A' arm of the trials acts as the common comparator and is called 'control'.

(b) Two (or more) rows per trial, one per trial arm (see *Table 25*)

Each row states the number of events (event) and participants (sample) and study arm (arm). This data structure is the standard format for fitting logistic and mixed models.

In Stata a data set of format (a) can be changed into format (b) using the following commands:

```
rename event_c event0
rename event_t event1
rename sample_c sample0
rename sample_t sample1
reshape long event sample, i(trial)
j(treat)
generate arm="C"
replace arm="A" if treat==1 & compb==0
replace arm="B" if treat==1 & compb==1
```

(c) Four (or more) rows per trial, two rows per trial arm (see *Table 26*)

Each row states the number in each outcome category (n) and indicator variables for whether they suffered a stroke or not (stroke) and study arm (arm). This data structure is only required for the models fitted in the Stata mixed models command gllamm.

TABLE 25 *Example data set for full analysis*

ID	Trial	Event	Sample	Arm
1	AFASAK	19	336	A
2	AFASAK	9	355	B
3	AFASAK	16	336	C
4	SPAK	19	211	A
5	SPAK	8	210	B
…	…	…	…	…
35	MWNAF	1	153	A
36	MWNAF	5	150	D

TABLE 26 *Example data set for full analysis in gllamm*

ID	Trial	n	Stroke	Arm
1	AFASAK	19	1	A
2	AFASAK	317	0	A
3	AFASAK	9	1	B
4	AFASAK	346	0	B
5	AFASAK	16	1	C
6	AFASAK	320	0	C
7	SPAK	19	1	A
8	SPAK	192	0	A
9	SPAK	8	1	B
10	SPAK	202	0	B
…	…	…		…
69	MWNAF	1	1	A
70	MWNAF	152	0	A
71	MWNAF	5	1	D
72	MWNAF	145	0	D

In Stata a data set of format (b) can be changed into format (c) using the following commands:

```
rename event n1
gene n0=sample-n1
reshape long event sample, i(trial arm)
j(stroke)
```

TABLE 27 *Example data set for full analysis in xtlogit*

ID	Trial	Stroke	Arm
1	AFASAK	1	A
2	AFASAK	1	A
3	AFASAK	1	A
…	…		…
19	AFASAK	1	A
20	AFASAK	0	A
21	AFASAK	0	A
…	…		…
8139	MWNAF	0	D
8140	MWNAF	0	D

(d) One row per participant, n rows per trial (see *Table 27*)

Each row states whether the participant suffered a stroke or not (stroke) and their study arm (arm). This data structure is only required for the models fitted in the Stata xtlogit.

In Stata a data set of format (c) can be changed into format (d) using the following commands:

```
expand n
drop n
```

Indirect comparisons

Naive method

The naive analysis is based on summing the events and participants in the B and C arms of the seven component trials, and computing an odds ratio and confidence interval as if the data had arisen in a single study.

Across the three B trial arms 16 out of 680 participants experienced a stroke. Across the four C trial arms 69 out of 506 participants experienced a stroke. This comparison yields an odds ratio of 0.15 (95% CI 0.09 to 0.27), and is highly statistically significant ($z = -6.61$, $p < 0.00001$).

Adjusted indirect comparisons and meta-regression

Adjusted indirect comparisons are undertaken by performing separate meta-analyses on the data sets of trials of AvB and AvC. One approach is to use the Stata meta-analysis command metan[103] for each meta-analysis (lines 5 and 9), and store the values of the log odds ratio (lines 6 and 10) and standard error (lines 7 and 11) from each separate meta-analysis. The odds ratio for indirect comparison is estimated as the exponential of the difference in log odds ratios from the two meta-analyses (lines 13 and 14). The standard error of the indirect comparison is estimated from the square root of the sum of the squared standard errors (line 15).

The code below assumes that the data are in format (a):

```
  * STATA CODE for indirect comparison by
    adjusted indirect comparison
1 * compute values of the four cells for
  each trial
2 gene noevent_t=sample_t-event_t
3 gene noevent_c=sample_c-event_c

4 * meta-analysis of trials of BvA (combp
  equal to 1)
5 metan event_t noevent_t event_c
  noevent_c if compb==1, or fixedi
  nograph
6 local logor1=log($S_1)
7 local se1=$S_2

8 * meta-analysis of trials of CvA (combp
  equal to 0)
9 metan event_t noevent_t event_c
  noevent_c if compb==0, or fixedi
  nograph
10 local logor2=log($S_1)
11 local se2=$S_2

12 * computation of log OR, OR and se for
   indirect comparison
13 local logor_aic=`logor1'-`logor2'
14 local or_aic=exp(`logor_aic')
15 local se_aic=sqrt(`se1'^2+`se2'^2)

16 * computation of confidence intervals,
   z-value and P-value
17 local ll_aic=exp(`logor_aic'-
   (1.96*`se_aic'))
18 local
   ul_aic=exp(`logor_aic'+(1.96*`se_aic')
   )
19 local z_aic=`logor_aic'/`se_aic'
20 if `z_aic'>0 local p_aic=2*(1-
   norm(`z_aic'))
21 if `z_aic'<=0 local
   p_aic=2*norm(`z_aic')
```

The above analysis produces inverse variance estimates. Alternative meta-analysis models are obtained by changing the inverse variance fixed effect option (fixedi) on the metan command lines (lines 5 and 9) to indicate Mantel–Haenszel fixed effects (fixed) or DerSimonian and Laird random

TABLE 28

Approach	Method	OR	95% CI	z/t-Value	p-Value
AIC	Inverse variance	0.43	(0.22 to 0.87)	$z = -2.36$	0.0184
AIC	Mantel–Haenszel	0.42	(0.21 to 0.83)	$z = -2.48$	0.0131
AIC	DerSimonian and Laird	0.43	(0.21 to 0.89)	$z = -2.28$	0.0225
Meta-regression	Weighted linear regression	0.43	(0.23 to 0.82)	$t = -3.37$	0.0198
Meta-regression	Random effects meta-regression	0.43	(0.22 to 0.87)	$z = -2.36$	0.0184

effects (randomi) models. Risk ratio models can be obtained by changing the summary statistic from odds ratio (or) to the risk ratio (rr). Risk difference estimates are produced using the risk difference (rd) option. If risk differences are pooled, use of a logarithmic scale is not required, requiring alterations to lines 6, 10, 14, 17 and 18 of the above code. Continuous outcomes can also be pooled using the metan command.

Meta-regression can also be used to estimate the difference between the groups. To undertake meta-regression, estimates of the log odds ratio and standard error are computed for each study. Using the values stored by metan provides a short-cut for doing this (lines 5–7). A weighted regression model is then fitted with the log odds ratio as the outcome and the comparison (compb) as the predictor. The model weights each study according to the inverse variance (line 8). A simple approach to fit the model involves weighted linear regression (line 10). A more appropriate method is to use random effects meta-regression (line 12) that allows for unexplained variability between groups, as provided by the metareg command.[78]

```
   * STATA CODE for indirect comparison
   by meta-regression
 1 * compute values of the four cells for
   each trial
 2 gene noevent_t=sample_t-event_t
 3 gene noevent_c=sample_c-event_c

 4 * meta-analysis of all trials to
   compute logOR and se
 5 metan event_t noevent_t event_c
   noevent_c, or fixedi nograph notable
 6 local logor=log($S_1)
 7 local se=$S_2
 8 local wt=1/`se'^2

 9 * weighted linear regression
10 regress logor compb [aw=wt]

11 * random effects meta-regression
12 metareg logor compb, wsse(se)
```

The three AIC models (inverse variance, Mantel–Haeszel, and DerSimonian and Laird) and two regression models (fixed and random) give similar estimates of the indirect comparison (*Table 28*), but all are very different from the flawed naive method.

The AIC DerSimonian and Laird and random effects meta-regression use different approaches to estimate the unexplained between study variation, τ^2. The DerSimonian and Laird method estimates separate τ^2 values for comparisons of AvC ($\tau^2 = 0$) and AvB ($\tau^2 = 0.023$). The meta-regression model estimates a single τ^2 to describe the residual heterogeneity among all trials having accounted for the difference in estimates between trials comparing AvB and trials comparing AvC ($\tau^2 = 0$). The meta-regression approach is likely to be more efficient and precise as it involves one fewer parameter and estimates τ^2 using data from all trials. The alternative DerSimonian and Laird approach may be more appropriate when there are reasons to assume different τ^2 values in the two component analyses.

Generalised linear models

A logistic regression model can be fitted to data in format (b). The model includes two zero–one indicator variables, armA and armB, to indicate study arm (lines 1–5). Arm C is designated the baseline category with which comparisons are made. An estimate of the indirect comparison is obtained from the parameter estimate for armB (comparison of B with baseline C). To obtain an analysis that is stratified by trial, indicator variables for trial are included to allow each trial to have a different control group risk (lines 7–8). In Stata the command blogit is used to fit a logistic regression model indicating the number of outcomes (event), the number of participants (people) and the dependent variables (armC, armA and trial):

```
* STATA CODE for indirect comparison by
logistic regression
```

```
1 * generate indicator variables for armA
  and armB
2 gene armA=0
3 gene armB =0
4 replace armA=1 if arm=="A"
5 replace armB=1 if arm=="B"

6 * fit a logistic regression model
7 * xi: automatically adds indicator
  variables for trial
8 xi:blogit event people armC armA
  i.trial, or
```

Alternatively, the Stata command xtlogit can be used for data in format (d). Again indicator variables are generated for study arms, which take values of 0 or –1 (lines 1–5). The i(trial) option combined with the fixed effect option fe (line 7) performs an analysis stratified by trial that produces the same results as the above logistic regression model.

The xtlogit command includes an option to perform random effects analysis, using the re option (line 9). This analysis replaces the trial indicator variables with a distribution of effects (making an assumption that the logit control group risks are a random sample from a normal distribution). It does not, as is usually desired in a meta-analytical random effects analysis, place the random effect on the treatment contrasts (see Chapter 4, section 'Classical methods using aggregate data', p. 19).

```
 * STATA CODE for indirect comparison
using xtlogit
1 * generate indicator variables for
  armA and armB
2 gene armA=0
3 gene armB =0
4 replace armA=-1 if arm=="A"
5 replace armB=-1 if arm=="B"

6 * fit a fixed effect logistic
  regression model
7 xtlogit stroke armC armA, i(trial)
  fe or

8 * fit a random effects logistic
  regression model
9 xtlogit stroke armC armA, i(trial)
  re or
```

Two standard software packages can be used for estimating a random effect for the treatment contrast for binomial data: SAS PROC NLMIXED and the gllamm command in Stata. The use of the SAS option is demonstrated here; this uses adaptive quadrature to identify maximum likelihood solutions, a more advanced and faster method than used by gllamm. WinBugs software,[81] using a fully Bayesian model specification, provides the greatest flexibility for fitting these models.

Two alternative models can be fitted using data in format (b) with indicator variables armA and armB indicating the treatment. In the first, trial effects are fitted as random (using the term trialr in lines 4 and 7) and a random effect for treatment contrasts is included assumed constant across treatment comparisons (using the term het in lines 4 and 7). The procedure requires appropriate starting values to be stated (line 3), which may be obtained from a fixed effect analysis.

```
* SAS code for indirect comparison using
  PROC NLMIXED
1 * random effects for trial and
  treatment contrasts
2 proc nlmixed data=indirect;
3   parms base=-2.3 a=0.2 b=-1
    s2trialr=0.7 s2het=0.1;
4   logitp= (base+trialr)+ armA*(a+het) +
    armB*(b+het);
5   p = exp(logitp)/(1+exp(logitp));
6   model stroke ~ binomial(n,p);
7   random trialr het ~
    normal([0,0],[s2trialr,0,s2het])
    subject=trial;
8 run;
```

This analysis estimates the trial effect variance to be 0.68, and the random effect for treatment contrasts to be 0.09. If desired, separate random effects can be estimated for the AvC and BvC treatment contrasts by specifying two separate heterogeneity parameters, although there are likely to be estimation problems unless there are many trials. This model requires the assumption that the logit baseline event rates are randomly sampled from a normal distribution. That assumption can be avoided by estimating a fixed effect for each trial while including a random effect for treatment contrast. This model requires estimation of many more parameters (additional parameters t3–t10 are estimated for each trial in lines 7 and 8) and is less likely to produce a stable solution. Again separate heterogeneity parameters could be included for each treatment contrast.

```
 * SAS code for indirect comparison
 using PROC NLMIXED
1 * fixed effect for trial, random
  effects for treatment contrasts
```

TABLE 29

Method	Random effects	OR	95% CI	z/t-Value	p-Value
Logistic regression (blogit, xtlogit fe)	None	0.42	(0.21 to 0.83)	$z = -2.49$	0.0129
Logistic regression (xtlogit re)	Trials	0.41	(0.22 to 0.78)	$z = -2.71$	0.007
Mixed model (NLMIXED)	Trials	0.38	(0.16 to 0.94)[a]	$t = -2.76$	0.03
Mixed model (NLMIXED)	Trials and treatments	0.38	(0.16 to 0.94)[a]	$t = -2.76$	0.04
Mixed model (NLMIXED)	Treatments	Unstable estimates (variance estimate is very small)			

[a] Treatment contrast calculated using *t* statistic.

```
 2 proc nlmixed data=indirect;
 3    parms a=0.2 b=-0.9
 4          t3=-3 t4=-3 t5=-2.7 t7=-1.3
            t8=-1.6 t9=-3.6 t10=-1.2
 5       s2het=1;
 6    logitp=a*(armA+het) + b*(armB+het)
+
 7       t3*trial3 + t4*trial4 +
         t5*trial5 +
 8    t7*trial7 +   t8*trial8 + t9*trial9
                   + t10*trial10;
 9    p = exp(logitp)/(1+exp(logitp));
10 model stroke ~ binomial(n,p);
11 random het ~ normal([0],[s2het])
   subject=trial;
12 run;
```

In the application to the example data set, the estimate of among-study heterogeneity is close to zero, but the estimates of the treatment contrasts are sensitive to the choice of starting value for the heterogeneity statistic (*Table 29*).

Combining indirect and direct comparisons

Trial 11 (SPAF II) is a two-armed trial of BvC. The estimated treatment effect in this trial is OR=0.90 (95% CI 0.45 to 1.79) (p=0.74). This section looks at analyses that combine this trial with the seven indirect trials.

Combining with adjusted indirect comparisons

A weighted combination of the results from the AIC and the direct comparison is computed as an inverse variance weighted average.

For the direct trial:

$$\text{lnOR} = \ln[(19/339)/(21/336)] = \ln(0.90) = -0.1090$$
$$\text{Var(lnOR)} = 1/19 + 1/339 + 1/21 + 1/336 = 0.1062$$

For the indirect comparison (from the inverse variance solution)

$$\text{lnOR} = -0.8358$$
$$\text{Var(lnOR)} = 0.1257$$

Inverse variance weights for the indirect and direct comparisons are 7.96 and 9.42, respectively. The weighted average and variance of the combination are thus:

$$\text{lnOR} = [(9.42 \times -0.1090) + (7.96 \times -0.8358)]/(7.96 + 9.42) = -0.4456$$
$$\text{Var(lnOR)} = 1/(7.96 + 9.42) = 0.0576$$

giving an overall estimate of OR = 0.64 (95% CI 0.40 to 1.02) (p = 0.06).

It is also possible to test whether there is a significant difference in the finding of the AIC and the direct comparison, by dividing the difference in the logOR (–0.1090 – –0.8358 = 0.7268) by the standard error of the difference [$\sqrt{(0.1062+0.1257)}$ = 0.4816], and comparing the resulting number (0.7268/0.4816 = 1.51) with a standard normal distribution (z) to obtain a p-value (p = 0.13).

It is possible to undertake the analyses combining any of the estimates using AICs or meta-regression models with the results of the direct trial. If there were more than one direct two-armed trial, results of a meta-analysis of the direct trials could be combined with the AIC.

TABLE 30

Approach	Method	OR	95% CI	z/t-Value	*p*-Value
AIC	Inverse variance	0.64	(0.40 to 1.02)	$z = -1.86$	0.0632
AIC	Mantel–Haenszel	0.63	(0.40 to 1.01)	$z = -1.94$	0.0530
AIC	DerSimonian and Laird	0.65	(0.40 to 1.04)	$z = -1.78$	0.0747
Meta-regression	Weighted linear regression	0.56	(0.38 to 0.83)	$t = -2.91$	0.0036
Meta-regression	Random effects meta-regression	0.64	(0.40 to 1.02)	$z = -1.86$	0.0632
Logistic regression	Fixed effect	0.62	(0.39 to 0.99)	$z = -2.00$	0.0460
Logistic regression	Random effect for trials	0.59	(0.37 to 0.93)	$z = -2.28$	0.0230
Mixed model	Random effect for trials	0.59	(0.34 to 1.03)	$t = -2.24$	0.0604
Mixed model	Random effect for trials and treatment	0.59	(0.33 to 1.05)	$t = -2.24$	0.0667
Mixed model	Random effect for treatment	Unstable estimates (variance estimate is very small)			

TABLE 31

Approach	Method	OR	95% CI	z/t-Value	*p*-Value
Logistic regression	Random effect for trials	0.57	(0.43 to 0.75)	$z = -3.99$	0.0001
Mixed model	Random effect for trials	0.53	(0.38 to 0.75)	$t = -4.09$	0.0022
Mixed model	Random effect for trials and treatment	0.53	(0.37 to 0.75)	$t = -4.09$	0.0027
Mixed model	Random effect for treatment	Unstable estimates (variance estimate is very small)			

TABLE 32

Approach	Method	OR	95% CI	z/t-Value	*p*-Value
Logistic regression	Random effect for trials	0.56	(0.43 to 0.73)	$z = -4.37$	<0.0001
Mixed model	Random effect for trials	0.56	(0.41 to 0.75)	$t = -4.15$	0.001
Mixed model	Random effect for trials and treatment	0.55	(0.41 to 0.75)	$t = -4.15$	0.001
Mixed model	Random effect for treatment	Unstable estimates (variance estimate is very small)			

Generalised linear models

The models described above can all incorporate data from the additional direct trial without changing the code other than inclusion of an additional indicator variable for the extra trial. The results for these models are given in *Table 30*.

Including three-armed AvBvC trials

Trials with more than two arms cannot be used in AICs without discarding some treatment arms. However, they can be included naturally in basic generalised linear models, again without adjustment to the code beyond addition of indicator variables for the additional trials. There are challenges in appropriately modelling random effects for multiarmed trials, as outlined by Higgins and Whitehead.[53] Appropriate modelling of random effects involves realising a different random effect for each comparison. SAS PROC NLMIXED assumes that all comparisons from the same trial involve the same realisation of the treatment comparison random effect, which is not ideal. Correct modelling can be undertaken using a fully Bayesian model in WinBUGS.[81]

Analysis of all trials and all treatments

AICs can be combined to include comparisons with alternative control groups only if all trials are two-armed trials. Here this is not the case, and incorporation of data from trials 12–15 can only be done using generalised linear models. In these situations additional indicator variables need to be added to the model for treatments D and E, as well as for the additional trials. The same issues arise in the correct modelling of random effects as discussed in the previous section.

Appendix 8

Identified meta-analyses providing sufficient data for both direct and indirect comparisons

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
1 ATC-I[17]	Patients with an increased risk of occlusive vascular disease (e.g. prior or acute CHD, stroke, peripheral vascular disease) (Vascular events)	High-dose aspirin vs medium-dose aspirin: High-dose aspirin vs control: Medium-dose aspirin vs control:	3 (1212/1213) 30 (13,667/11,565) 19 (25,376/25,417)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.96 (0.81 to 1.15) 1.08 (0.94 to 1.24) 1.64 (1.54 to 1.74)
2 ATC-I[17]	Same as above	Aspirin+dipyridamole vs aspirin: Aspirin+dipyridamole vs control: Aspirin vs control:	16 (2829/2840) 23 (4757/4694) 42 (40,013/37,538)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.01 (0.87 to 1.16) 0.91 (0.80 to 1.05) 1.12 (1.03 to 1.22)
3 ATC-I[17]	Same as above	Sulfinpyrazone vs aspirin: Sulfinpyrazone vs control: Aspirin vs control:	4 (507/656) 17 (2108/2135) 49 (41,656/38,799)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.17 (0.88 to 1.54) 1.02 (0.87 to 1.20) 1.32 (1.17 to 1.47)
4 ATC-I[17]	Same as above	Ticlopidine vs aspirin: Ticlopidine vs control: Aspirin vs control:	3 (1730/1741) 27 (2936/2955) 52 (42,248/39,230)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.71 (0.38 to 1.34) 0.90 (0.78 to 1.04) 0.91 (0.81 to 1.03)
5 ATC-II[16]	Patients with an increased risk of vascular occlusion (e.g. coronary or leg artery bypass grafting or angioplasty) (Vascular occlusion)	High-dose vs medium-dose aspirin: High-dose aspirin vs control: Medium-dose aspirin vs control:	1 (155/154) 8 (744/753) 4 (489/496)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.15 (0.76 to 1.74) 0.93 (0.58 to 1.48) 0.59 (0.46 to 0.77)
6 ATC-II[16]	Same as above	Aspirin+dipyridamole vs aspirin: Aspirin+dipyridamole vs control: Aspirin vs control:	10 (1264/1267) 14 (1371/1303) 7 (597/610)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.03 (0.84 to 1.27) 1.26 (0.85 to 1.86) 1.53 (1.21 to 1.92)
7 ATC-II[16]	Same as above	Sulfinpyrazone vs aspirin: Sulfinpyrazone vs control: Aspirin vs control:	2 (167/326) 5 (276/285) 12 (1233/1249)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.01 (0.71 to 1.45) 1.15 (0.73 to 1.80) 0.81 (0.58 to 1.13)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
8 ATC-II[16]	Same as above	Ticlopidine vs aspirin: Ticlopidine vs control: Aspirin vs control:	2 (41/41) 12 (546/542) 13 (1542/1402)	RR Direct estimate: Adjusted indirect: Naive Indirect:	 1.16 (0.43 to 3.16) 1.12 (0.80 to 1.56) 0.75 (0.59 to 0.96)
9 ATC-II[16]	Same as above	Aspirin+dipyridamole vs sulfinpyrazone: Aspirin+dipyridamole vs control: Sulfinpyrazone vs control :	1(162/148) 19 (2004/1942) 5 (276/285)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.94 (0.62 to 1.43) 1.06 (0.69 to 1.65) 1.54 (1.13 to 2.12)
10 ATC-III[18]	Surgical and high-risk medical patients (DVT)	Aspirin+dipyridamole vs aspirin: Aspirin+dipyridamole vs control: Aspirin vs control:	9 (263/218) 11 (394/422) 9 (649/597)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.67 (0.51 to 0.89) 0.77 (0.44 to 1.33) 0.94 (0.73 to 1.20)
11 Bucher *et al.*[14]	HIV-infected patients (*Pneumocystis carinii* pneumonia)	TMP+SMX vs D+P or D: TMP+SMX vs AP: D+P or D vs AP:	8 (803/815) 9 (681/613) 5 (732/718)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.45 (0.22 to 0.91) 0.43 (0.25 to 0.75) 0.55 (0.35 to 0.87)
12 Cheng *et al.*[137]	Women attending services for emergency contraception (No. of pregnancies)	Levonorgestrel vs mifepristone: Levonorgestrel vs yuzpe: Mifepristone vs yuzpe:	1 (643/633) 2 (1386/1421) 2 (597/589)	RR Direct estimate: Adjusted indirect: Naive indirect:	 2.19 (1.00 to 4.77) 5.40 (0.53 to 54.82) 19.83 (1.21 to 326.03)
13 Chiba *et al.*[40]	Patients with GORD (Healing rate)	H_2RA vs PPI: H_2RA vs placebo: PPI vs placebo:	13 (731/884) 11 (1817/959) 2 (334/75)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.56 (0.48 to 0.66) 0.26 (0.14 to 0.48) 0.73 (0.68 to 0.78)
14 Collins *et al.*[138]	Patients with postoperative pain (No. of patients with >50% pain relief)	Ibuprofen 400 mg vs 200 mg: Ibuprofen 400 mg vs control: Ibuprofen 200 mg vs control:	5 (199/202) 26 (1407/1043) 3 (204/133)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.39 (1.08 to 1.79) 0.74 (0.27 to 2.02) 1.16 (1.00 to 1.34)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
15 Delaney *et al.*[139]	Patients with dyspepsia (Global assessment)	PPI vs H_2RA: PPI vs alginate/antacid: H_2RA vs alginate/antacid:	3 (633/634) 2 (595/591) 1 (119/136)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.64 (0.49 to 0.82) 0.73 (0.56 to 0.96) 0.86 (0.71 to 1.05)
16 Di Mario *et al.*[140]	Patients with previously untreated gastric ulcer (Endoscopic healing)	Cimetidine vs ranitidine: Cimeditine vs placebo: Ranitidine vs placebo:	5 (total 636) 13 (total 852) 8 (total 756)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.21 (0.88 to 1.67) 0.65 (0.35 to 1.20) Not available
17 Handoll *et al.*[141]	Patients undergoing surgery for hip fractures (DVT)	LMWH vs UFH: LMWH vs placebo: UFH vs placebo:	3 (136/111) 2 (104/110) 10 (407/409)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.91 (0.36 to 2.31) 1.05 (0.52 to 2.13) 0.68 (0.44 to 1.07)
18 Horn and Limburg[142]	Patients with acute ischemic stroke (Poor outcome: death or dependency in activities of daily living)	Mimodipine 240 mg vs 120 mg: Mimodipine 240 mg vs control: Mimodipine 120 mg vs control:	2 (340/341) 1 (73/69) 13 (2081/2106)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.07 (0.94 to 1.22) 0.97 (0.63 to 1.48) 1.06 (0.79 to 1.42)
19 Marshall and Irvine[21]	Patients with ulcerative colitis (Endoscopic remission)	5-ASA vs rectal corticosteroids: 5-ASA vs R. budesonide: Rectal corticosteroids vs rectal budesonide:	7 (total 682) 2 (total 154) 5 (total 463)	OR Direct estimate: Adjusted indirect: Naive indirect:	 0.53 (0.36 to 0.78) 0.92 (0.36 to 2.36) 1.13 (0.86 to 1.49)
20 McIntosh and Olliaro[143]	Patients with uncomplicated malaria (Parasite clearance at day 28)	Artemisinin vs artesunate: Artemisinin vs quinine: Artesunate vs quinine:	1 (20/19) 1 (27/22) 1 (47/39)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.82 (0.55 to 1.22) 0.70 (0.38 to 1.28) 0.42(0.26 to 0.66)
21 Moore *et al.*[24]	Patients with severe postoperative pain (>50% pain relief)	Paracetamol+codeine vs paracetamol: Paracetamol+codeine vs placebo: Paracetamol vs placebo:	10 (309/313) 5 (98/110) 7 (281/254)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.24 (1.01 to 1.54) 1.74 (0.59 to 5.18) 1.03 (0.78 to 1.35)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
22 Pagliaro (used by Higgins and Whitehead[53]) (Rate of first bleeding)	Patients with cirrhosis and oesophagogastric varices	β-Blockers vs sclerotherapy: β-Blockers vs control: Sclerotherapy vs control:	2 (111/115) 7 (378/394) 17 (723/736)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.53 (0.12 to 2.36) 1.00 (0.53 to 1.89) 0.69 (0.53 to 0.91)
23 Poynard *et al.*[26] (Sustained ALT response rate)	Patients with viral hepatitis C	Interferon (3 MU) 12 months vs 6 months: Interferon (3 MU) 12 months vs control: Interferon (3 MU) 6 months vs control:	4 (256/249) 5 (161/157) 6 (132/131)	RR Direct estimate: Adjusted indirect: Naive indirect:	 2.20 (1.52 to 3.17) 1.49 (0.35 to 6.31) 1.67 (1.15 to 2.42)
24 Rostom *et al.*[107] (Total endoscopic ulcers)	Patients who had taken NSAIDs for >3 weeks	PPI vs H_2RA: PPI vs placebo: H_2RA vs placebo:	1 (210/215) 3 (443/331) 6 (645/541)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.28 (0.15 to 0.51) 0.61 (0.40 to 0.93) 1.02 (0.72 to 1.44)
25 Silagy[144] (Smoking cessation rate)	Smokers	>1 visit vs 1 visit: >1 visit vs control: 1 visit vs control:	5 (733/521) 4 (1931/1529) 15 (7551/5826)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.51 (1.08 to 2.12) 1.51 (0.90 to 2.56) 2.34 (2.02 to 2.71)
26 Silagy *et al.*[145] (Smoking cessation rate)	Patients undergoing NRT	Nicotine patch 24 hours vs 16 hours Nicotine patch 25 hours vs control: Nicotine patch 16 hours vs control:	1 (51/55) 24 (5415/4304) 8 (4368/1737)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.70 (0.36 to 1.35) 0.82 (0.56 to 1.20) 1.07 (0.97 to 1.18)
27 Silagy *et al.*[145]	Same as above	NRT weaning vs no weaning: NRT weaning vs control: NRT no weaning vs control:	1 (68/56) 24 (7571/4598) 5 (701/648)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.97 (0.68 to 1.38) 0.75 (0.53 to 1.06) 0.88 (0.74 to 1.05)
28 Soo *et al.*[108]	Patients with non-ulcer dyspepsia (Global symptom assessment)	H_2RA vs sucralfate: H_2RA vs placebo: Sucralfate vs placebo:	1 (47/53) 8 (607/618) 2 (129/117)	RR Direct estimate: Adjusted indirect: Naive indirect:	 2.74 (1.25 to 6.02) 0.99 (0.47 to 2.08) 1.24 (0.91 to 1.70)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
29 Soo *et al.*[108]	Same as above	Prokinetics vs H_2RA: Prokinetics vs placebo: Sucralfate vs placebo:	2 (208/215) 10 (326/264) 7 (496/508)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.54 (0.22 to 1.33) 0.66 (0.41 to 1.05) 1.04 (0.84 to 1.28)
30 Trinadade and Menon[146]	Patients with major depression (No. of dropouts)	Fluoxetine vs paroxetine: Fluoxetine vs controls: Paroxetine vs controls:	5 (327/328) 40 (2500/2234) 33 (1471/1423)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.00 (0.73 to 1.37) 1.00 (0.86 to 1.17) 1.24 (1.13 to 1.35)
31 Trindade and Menon[146]	Same as above	Fluoxetine vs fluvoxamine: Fluoxetine vs controls: Fluvoxamine vs controls:	1 (49/51) 45 (2909/2378) 41 (1168/1166)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.78 (0.18 to 3.31) 0.89 (0.77 to 1.02) 1.14 (1.03 to 1.26)
32 Trindade and Menon[146]	Same as above	Paroxetine vs fluvoxamine: Paroxetine vs controls: Fluvoxamine vs controls:	1 (56/64) 34 (1335/1288) 35 (1025/1019)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.80 (0.47 to 1.35) 0.77 (0.63 to 0.93) 0.92 (0.81 to 1.03)
33 Trindade and Menon[146]	Same as above	Sertraline vs fluvoxamine: Sertraline vs controls: Fluvoxamine vs controls:	1 (48/49) 9 (816/759) 23 (685/651)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.40 (0.18 to 0.86) 0.81 (0.60 to 1.10) 0.94 (0.81 to 1.09)
34 Trindede and Menon[146]	Same as above	Fluoxetine vs sertraline: Fluoxetine vs controls: Sertraline vs controls:	3 (266/273) 35 (2502/1967) 9 (816/759)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.90 (0.38 to 2.14) 0.88 (0.70 to 1.11) 1.10 (0.98 to 1.23)
35 van Pinxteren *et al.*[109]	Patients with GORD-like symptoms (Heartburn remission)	PPI vs H_2RA: PPI vs placebo: H_2RA vs placebo:	3 (1228/664) 1 (161/159) 2 (511/502)	RR Direct estimate: Adjusted indirect: Naive indirect:	 0.67 (0.57 to 0.80) 0.45 (0.31 to 0.66) 0.53 (0.40 to 0.71)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
36 Zhang and Li Wan Po[147]	Patients with dysmenorrhoea (No. of patients with at least moderate pain relief)	Naproxen vs ibuprofen: Naproxen vs placebo: Ibuprofen vs placebo:	3 (122/113) 17 (904/877) 10 (345/346)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.08 (0.79 to 1.48) 1.40 (0.94 to 2.09) 0.82 (0.75 to 0.89)
37 Zhang and Li Wan Po[147]	Same as above	Naproxen vs mefenamic acid: Naproxen vs placebo: Mefenamic vs placebo:	1 (24/20) 18 (942/916) 4 (307/307)	RR Direct estimate: Adjusted indirect: Naive indirect:	 2.40 (1.39 to 4.13) 1.53 (1.11 to 2.12) 0.88 (0.80 to 0.97)
38 Zhang and Li Wan Po[147]	Same as above	Naproxen vs aspirin: Naproxen vs placebo: Aspirin vs placebo:	1 (32/32) 17 (890/872) 5 (220/223)	RR Direct estimate: Adjusted indirect: Naive indirect:	 2.29 (1.16 to 4.52) 2.45 (1.65 to 3.64) 1.83 (1.49 to 2.23)
39 Zhang and Li Wan Po[147]	Same as above	Ibuprofen vs aspirin: Ibuprofen vs placebo: Aspirin vs placebo:	1 (43/43) 10 (322/327) 5 (187/184)	RR Direct estimate: Adjusted indirect: Naive indirect:	 1.90 (1.30 to 2.77) 1.80 (1.12 to 2.89) 2.32 (1.84 to 2.93)
40 Ausejo[148]	Children with croup (Improvement in croup severity score)	Budesonide vs dexamethasone: Budesonide vs placebo: Dexamethasone vs placebo:	1 (65/69) 5 (166/161) 8 (365/374)	Standardised mean difference Direct estimate: Adjusted indirect: Naive indirect:	 0.09 (–0.25 to 0.43) 0.32 (–0.52 to 1.16) 0.20 (0.02 to 0.38)
41 Li Wan Po and Zhang[20]	Patients with postsurgical pain (Sum of difference in pain intensity)	Paracetamol+dexamethasone vs paracetamol: Paracetamol+dexamethasone vs placebo: Paracetamol vs placebo:	3 (103/99) 5 (181/178) 14 (558/534)	Mean difference Direct estimate: Adjusted indirect: Naive indirect:	 1.22 (0.00 to 2.45) 0.51 (–0.43 to 1.45) 0.13 (–0.61 to 0.88)
42 Packer *et al.*[69]	Patients with heart failure (Changes in left ventricular ejection fraction)	Carvedilol vs metoprolol: Carvedilol vs placebo: Metoprolol vs placebo:	4 (123/125) 9 (534/668) 6 (376/408)	Mean difference Direct estimate: Adjusted indirect: Naive indirect:	 0.029 (0.007 to 0.051) 0.027 (0.013 to 0.041) 0.016 (0.012 to 0.0204)

continued

TABLE 33 *Identified meta-analyses with sufficient data for both direct and indirect comparisons (cont'd)*

Meta-analysis	Patients (outcome)	Interventions compared: no. of trials (patients)		Relative efficacy (95% CI)	
43 Sauriol *et al.*[149]	Patients with schizophrenia (Changes in brief psychiatric rating scale)	Olanzapine vs risperidone: Olanzapine vs placebo: Risperidone vs placebo:	1 (172/167) 3 (1620/786) 8 (1044/416)	Mean difference Direct estimate: Adjusted indirect: Naive indirect:	 1.80 (–1.43 to 5.03) 1.33 (–0.63 to 3.29) 1.07 (–0.28 to 2.42)
44 Zhang and Li Wan Po[147]	Patients with surgical pain (Sum of difference in pain intensity)	Paracetamol+codeine vs paracetamol: Paracetamol+codeine vs placebo: Paracetamol vs placebo:	13 (449/448) 12 (NA) 31 (NA)	Mean difference Direct estimate: Adjusted indirect: Naive indirect:	 6.97 (3.56 to 10.37) –1.16 (–6.95 to 4.64) –9.89 (–11.65 to –8.13)

AP, aerolised pentamidine; D, dapsone; NA, not available; NRT, nicotine replacement therapy; P, pyrimethamine; SMX, sulfamethoxazole; TMP, trimethoprim.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology & Therapeutics,
University of Liverpool

Professor Bruce Campbell,
Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Dr Edmund Jessop, Medical Advisor, National Specialist, Commissioning Advisory Group (NSCAG), Department of Health, London

Professor Jon Nicholl, Director, Medical Care Research Unit, University of Sheffield, School of Health and Related Research

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology & Therapeutics,
University of Liverpool

Chair,
Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield, School of Health and Related Research

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine

Dr Jeffrey Aronson
Reader in Clinical Pharmacology, Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford

Professor Deborah Ashby,
Professor of Medical Statistics, Department of Environmental and Preventative Medicine, Queen Mary University of London

Professor Ann Bowling,
Professor of Health Services Research, Primary Care and Population Studies,
University College London

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford

Professor John Cairns, Professor of Health Economics, Public Health Policy, London School of Hygiene and Tropical Medicine, London

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York

Mr Jonathan Deeks,
Senior Medical Statistician,
Centre for Statistics in Medicine, University of Oxford

Dr Andrew Farmer, Senior Lecturer in General Practice,
Department of Primary Health Care,
University of Oxford

Professor Fiona J Gilbert,
Professor of Radiology,
Department of Radiology,
University of Aberdeen

Professor Adrian Grant,
Director, Health Services Research Unit, University of Aberdeen

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry,
University of Cambridge

Professor Sallie Lamb,
Professor of Rehabilitation,
Centre for Primary Health Care,
University of Warwick

Professor Stuart Logan,
Director of Health & Social Care Research, The Peninsula Medical School,
Universities of Exeter & Plymouth

Dr Linda Patterson,
Consultant Physician,
Department of Medicine,
Burnley General Hospital

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine

Professor Mark Sculpher,
Professor of Health Economics,
Centre for Health Economics,
Institute for Research in the Social Services, University of York

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre,
Birmingham

Ms Kate Thomas,
Deputy Director,
Medical Care Research Unit,
University of Sheffield

Ms Sue Ziebland,
Research Director, DIPEx,
Department of Primary Health Care, University of Oxford,
Institute of Health Sciences

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Lay Member, Bolton

Professor Max Bachmann Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant Paediatrician/ Hon. Senior Lecturer, Population Health Unit, Great Ormond St. Hospital, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medicines & Healthcare Products Regulatory Agency, London

Professor William Rosenberg, Professor of Hepatology, Liver Research Group, University of Southampton

Dr Susan Schonfield, Consultant in Public Health, Specialised Services Commissioning North West London, Hillingdon Primary Care Trust

Dr Phil Shackley, Senior Lecturer in Health Economics, School of Population and Health Sciences, University of Newcastle upon Tyne

Dr Margaret Somerville, PMS Public Health Lead, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Associate Dean for Education, Head of Department of Obstetrics and Gynaecology, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Chair Division A, The John Radcliffe Hospital, Oxford Radcliffe Hospitals NHS Trust

Professor Tony Avery, Head of Division of Primary Care, School of Community Health Services, Division of General Practice, University of Nottingham

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Consultant in Pharmaceutical Public Health, National Public Health Service for Wales, Cardiff

Mrs Sharon Hart, Head of DTB Publications, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, South Gloucestershire Primary Care Trust

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, The Royal Marsden Hospital, Sutton

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Aileen Clarke,
Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, London

Dr Matthew Cooke, Reader in A&E/Department of Health Advisor in A&E, Warwick Emergency Care and Rehabilitation, University of Warwick

Dr Carl E Counsell, Clinical Senior Lecturer in Neurology, Department of Medicine and Therapeutics, University of Aberdeen

Ms Amelia Curwen, Executive Director of Policy, Services and Research, Asthma UK, London

Professor Gene Feder, Professor of Primary Care R&D, Department of General Practice and Primary Care, Barts & the London, Queen Mary's School of Medicine and Dentistry, London

Professor Paul Gregg,
Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Bec Hanley, Co-Director, TwoCan Associates, Hurstpierpoint

Ms Maryann L Hardy,
Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich,
Director of Clinical R&D, Academic Department of Radiology, The Institute of Cancer Research, London

Dr Simon de Lusignan,
Senior Lecturer,
Primary Care Informatics,
Department of Community Health Sciences,
St George's Hospital Medical School, London

Professor Neil McIntosh,
Edward Clark Professor of Child Life & Health,
Department of Child Life & Health, University of Edinburgh

Professor James Neilson,
Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford,
Consultant Physician,
Directorate of Medical Services,
North Bristol NHS Trust

Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

Dr Vimal Sharma, Consultant Psychiatrist/Hon. Senior Lecturer, Mental Health Resource Centre, Cheshire and Wirral Partnership NHS Trust, Wallasey

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh,
Professor of Public Health,
Department of Public Health,
University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman,
Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond,
Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Professor Iain T Cameron,
Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark,
Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Clifford,
Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson,
Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Mr John Dunning,
Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Services, West Middlesex University Hospital, Isleworth

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Alastair Gray,
Professor of Health Economics, Department of Public Health, University of Oxford

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Dr Duncan Keeley,
General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping,
Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay,
Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Julian Little,
Professor of Human Genome Epidemiology, Department of Epidemiology & Community Medicine, University of Ottawa

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Sue Moss, Associate Director,
Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Mrs Julietta Patnick,
Director, NHS Cancer Screening Programmes, Sheffield

Professor Tim Peters,
Professor of Primary Care Health Services Research, Academic Unit of Primary Health Care, University of Bristol

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Professor Peter Sandercock,
Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Dr Ken Stein,
Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter

Professor Sarah Stewart-Brown,
Professor of Public Health, University of Warwick, Division of Health in the Community Warwick Medical School, LWMS, Coventry

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

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

ISSN 1366-5278