Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies

J Savović, HE Jones, DG Altman, RJ Harris,
P Jűni, J Pildal, B Als-Nielsen, EM Balk,
C Gluud, LL Gluud, JPA Ioannidis, KF Schulz,
R Beynon, N Welton, L Wood, D Moher,
JJ Deeks and JAC Sterne

HTA

How to obtain copies of this and other HTA programme reports

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues 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 issue and for the rest of the world £3 per issue.

How to order:

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

Additionally the HTA website allows you to either print out your order or download a blank order form.

Contact details are as follows:

Synergie UK (HTA Department)
Digital House, The Loddon Centre
Wade Road
Basingstoke
Hants RG24 8QW

Email: orders@hta.ac.uk

Tel: 0845 812 4000 – ask for 'HTA Payment Services'
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put 'HTA Order' on the fax header

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

Paying by credit card
You can order using your credit card by phone, fax or post.

Subscriptions

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

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

Please use the form on the HTA website (www.hta.ac.uk/htacd/index.shtml). *HTA on DVD* is currently free of charge worldwide.

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

Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies

J Savović,[1] HE Jones,[1] DG Altman,[2] RJ Harris,[3] P Jűni,[4] J Pildal,[5] B Als-Nielsen,[6] EM Balk,[7] C Gluud,[8] LL Gluud,[9] JPA Ioannidis,[10] KF Schulz,[11] R Beynon,[1] N Welton,[1] L Wood,[12] D Moher,[13] JJ Deeks[14] and JAC Sterne[1]*

[1]School of Social and Community Medicine, University of Bristol, Bristol, UK
[2]Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, UK
[3]Centre for Infections, Health Protection Agency, Colindale, UK
[4]Division of Clinical Epidemiology and Biostatistics, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; CTU Bern, Bern University Hospital, Bern, Switzerland
[5]The Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
[6]Copenhagen Trial Unit, Centre for Clinical Intervention Research and Department of Paediatrics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[7]Center for Clinical Evidence Synthesis, Tufts Clinical and Translational Science Institute, Tufts Medical Center, Boston, MA, USA
[8]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[9]Department of Internal Medicine, Copenhagen University Hospital Gentofte, Copenhagen, Denmark
[10]Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA
[11]Quantitative Sciences, Family Health International, Research Triangle Park, Durham, NC, USA
[12]School Food Trust, Sheffield, UK
[13]Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[14]Public Health Epidemiology and Biostatistics, School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

*Corresponding author

This report should be referenced as follows:

Savović J, Jones HE, Altman DG, Harris RJ, Jűni P, Pildal J, *et al*. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012;**16**(35).

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

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work 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.

Abstract

Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies

J Savović,[1] HE Jones,[1] DG Altman,[2] RJ Harris,[3] P Jűni,[4] J Pildal,[5] B Als-Nielsen,[6] EM Balk,[7] C Gluud,[8] LL Gluud,[9] JPA Ioannidis,[10] KF Schulz,[11] R Beynon,[1] N Welton,[1] L Wood,[12] D Moher,[13] JJ Deeks[14] and JAC Sterne[1]*

[1]School of Social and Community Medicine, University of Bristol, Bristol, UK
[2]Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, UK
[3]Centre for Infections, Health Protection Agency, Colindale, UK
[4]Division of Clinical Epidemiology and Biostatistics, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; CTU Bern, Bern University Hospital, Bern, Switzerland
[5]The Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
[6]Copenhagen Trial Unit, Centre for Clinical Intervention Research and Department of Paediatrics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[7]Center for Clinical Evidence Synthesis, Tufts Clinical and Translational Science Institute, Tufts Medical Center, Boston, MA, USA
[8]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[9]Department of Internal Medicine, Copenhagen University Hospital Gentofte, Copenhagen, Denmark
[10]Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA
[11]Quantitative Sciences, Family Health International, Research Triangle Park, Durham, NC, USA
[12]School Food Trust, Sheffield, UK
[13]Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[14]Public Health Epidemiology and Biostatistics, School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

*Corresponding author

Background: The design of randomised controlled trials (RCTs) should incorporate characteristics (such as concealment of randomised allocation and blinding of participants and personnel) that avoid biases resulting from lack of comparability of the intervention and control groups. Empirical evidence suggests that the absence of such characteristics leads to biased intervention effect estimates, but the findings of different studies are not consistent.
Objectives: To examine the influence of unclear or inadequate random sequence generation and allocation concealment, and unclear or absent double blinding, on intervention effect estimates and between-trial heterogeneity, and whether or not these influences vary with type of clinical area, intervention, comparison and outcome measure.

Data sources and methods: Data were combined from seven contributing meta-epidemiological studies (collections of meta-analyses in which trial characteristics are assessed and results recorded). The resulting database was used to identify and remove overlapping meta-analyses. Outcomes were coded such that odds ratios < 1 correspond to beneficial intervention effects. Outcome measures were classified as mortality, other objective or subjective. We examined agreement between assessments of trial characteristics in trials assessed in more than one contributing study. We used hierarchical Bayesian bias models to estimate the effect of trial characteristics on average bias [quantified as ratios of odds ratios (RORs) with 95% credible intervals (CrIs) comparing trials with and without a characteristic] and in increasing between-trial heterogeneity.

Results: The analysis data set contained 1973 trials included in 234 meta-analyses. Median kappa statistics for agreement between assessments of trial characteristics were: sequence generation 0.60, allocation concealment 0.58 and blinding 0.87. Intervention effect estimates were exaggerated by an average 11% in trials with inadequate or unclear (compared with adequate) sequence generation (ROR 0.89, 95% CrI 0.82 to 0.96); between-trial heterogeneity was higher among such trials. Bias associated with inadequate or unclear sequence generation was greatest for subjective outcomes (ROR 0.83, 95% CrI 0.74 to 0.94) and the increase in heterogeneity was greatest for such outcomes [standard deviation (SD) 0.20, 95% CrI 0.03 to 0.32]. The effect of inadequate or unclear (compared with adequate) allocation concealment was greatest among meta-analyses with a subjectively assessed outcome intervention effect (ROR 0.85, 95% CrI 0.75 to 0.95), and the increase in between-trial heterogeneity was also greatest for such outcomes (SD 0.20, 95% CrI 0.02 to 0.33). Lack of, or unclear, double blinding (compared with double blinding) was associated with an average 13% exaggeration of intervention effects (ROR 0.87, 95% CrI 0.79 to 0.96), and between-trial heterogeneity was increased for such studies (SD 0.14, 95% CrI 0.02 to 0.30). Average bias (ROR 0.78, 95% CrI 0.65 to 0.92) and between-trial heterogeneity (SD 0.37, 95% CrI 0.19 to 0.53) were greatest for meta-analyses assessing subjective outcomes. Among meta-analyses with subjectively assessed outcomes, the effect of lack of blinding appeared greater than the effect of inadequate or unclear sequence generation or allocation concealment.

Conclusions: Bias associated with specific reported study design characteristics leads to exaggeration of beneficial intervention effect estimates and increases in between-trial heterogeneity. For each of the three characteristics assessed, these effects were greatest for subjectively assessed outcomes. Assessments of the risk of bias in RCTs should account for these findings. Further research is needed to understand the effects of attrition bias, as well as the relative importance of blinding of patients, care-givers and outcome assessors, and thus separate the effects of performance and detection bias.

Funding: National Institute for Health Research Health Technology Assessment programme.

Contents

List of abbreviations

BRANDO	Bias in Randomised and Observational studies (study acronym)
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CrI	credible interval
ID	unique identifier/identification number in the database
IQR	interquartile range
OR	odds ratio
PMID	unique PubMed™ identifier
PPr	posterior probability
RCT	randomised controlled (clinical) trial
ROR	ratio of odds ratios

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 in the notes at the end of the table.

Executive summary

Background

Systematic reviews of randomised controlled trials (RCTs) provide the best evidence for clinical and policy decision-making about medical interventions. The design of RCTs should incorporate characteristics (such as concealment of randomised allocation and blinding of participants and personnel) that avoid biases resulting from lack of comparability of the intervention and control groups. Empirical evidence suggests that absence of such characteristics, as reported in trial publications, leads to biased estimates of intervention effects. Collections of meta-analyses assembled in meta-epidemiological studies are used to study associations of study design characteristics with intervention effect estimates, but findings vary between studies.

Objectives

- To combine data from contributing meta-epidemiological studies into a single database, and derive a harmonised data set in which overlap between meta-analyses was removed.
- To examine agreement in the assessment of reported study design characteristics in the subset of trials that was assessed in two or more contributing meta-epidemiological studies.
- To examine the influence of inadequate or unclear (compared with adequate) random sequence generation and allocation concealment, and absent or unclear double blinding (compared with double blinding), on intervention effect estimates and between-trial heterogeneity.
- To examine whether or not these influences vary with the type of clinical area, intervention, comparison and outcome measure.
- To examine the effects of combinations of characteristics, and to estimate adjusted effects using multivariable models.
- To explore the implications of these findings for downweighting of trials whose characteristics are associated with bias in future meta-analyses.

Methods

We combined data from 10 contributing meta-epidemiological studies into a single database containing 427 reviews, 454 meta-analyses and 4874 trial results, whose design allowed trials to be contained in different meta-analyses, multiple meta-analyses in systematic reviews, overlapping meta-analyses between systematic reviews and multiple references to the same trial or review. Unique identifiers were used to identify sets of meta-analyses with overlapping trials: 258 meta-analyses were unique, whereas for 196 at least one trial overlapped with another meta-analysis. Overlapping meta-analyses and trials were removed according to a pre-specified protocol. The final database contained 363 meta-analyses and 3477 unique trial results. Overlapping trials were used to estimate kappa statistics for agreement between assessments of reported study design characteristics.

Information on outcome measures, interventions and comparisons was extracted from the included systematic reviews. Interventions were classified as pharmacological; surgical; psychosocial and behavioural; and all other interventions. Comparison interventions were classified as inactive (placebo, no intervention, standard care) or active. Outcome measures were grouped as all-cause mortality; other objectively assessed; and subjectively assessed.

The main analyses excluded 87 meta-analyses (1093 trials) from four contributing studies that did not collect data on both study design characteristics and outcome events; 36 meta-analyses (300 trials) in which it was not possible to classify one intervention as experimental and the other as control; one meta-analysis (four trials) that had a continuous outcome measure; 45 trials with missing outcome data; and 50 trials in which either no or all subjects experienced the outcome event.

Statistical methods

Intervention effects were modelled as log-odds ratios; odds ratios < 1 corresponded to beneficial intervention effects. We fitted Bayesian hierarchical bias models that allowed for random intervention effects within meta-analyses, with meta-analysis-specific mean μ and between-trial variance τ^2. Three effects of study design characteristics were modelled. First, mean intervention effects among trials reported as having a particular study design characteristic may differ from those among trials without that characteristic: estimated mean differences were exponentiated and reported as ratios of odds ratios (RORs). Second, variation in bias between trials within meta-analyses was quantified by standard deviation κ; κ^2 corresponds to the average increase in between-trial heterogeneity in trials with a specified study design characteristic. Third, variation in mean bias between meta-analyses was quantified by between-meta-analysis standard deviation φ. We derived 95% credible intervals (CrIs) for each parameter. Bias models were fitted using WinBUGS Version 1.4 (MRC Biostatistics Unit, Cambridge, UK), using vague prior distributions. The prior for variance parameters found to give the best overall performance was a modified Inverse Gamma(0.001, 0.001) prior with increased weight on small values. For location parameters (overall mean bias, baseline response rates, treatment effects), Normal(0, 1000) priors were assumed.

We first conducted univariable analyses for each bias domain separately using all informative meta-analyses (meta-analyses containing trials with and without the characteristic of interest) for each characteristic. The primary analysis used dichotomised variables for each characteristic (inadequate or unclear compared with adequate sequence generation and allocation concealment, and not double blind or unclearly blinded compared with double bind). Analyses were stratified according to type of outcome measure (all-cause mortality, other objectively assessed and subjectively assessed). Further univariable analyses (examining the influence of combinations of design characteristics) and multivariable analyses were conducted using two data subsets: trials with information on all three characteristics and trials with information on both allocation concealment and blinding.

Results

There was good agreement between assessments of reported study design characteristics carried out in the different studies. For sequence generation (two comparisons), the percentages of studies in which the assessments were in agreement were 81% and 82% and kappa statistics were 0.56 and 0.64. For allocation concealment (12 comparisons), percentage agreement varied between 52% and 100% and kappa statistics between 0.19 and 1.00 (median 0.58). Assessments were most reliable for blinding (nine comparisons): percentage agreement ranged from 80% to 100% (in four comparisons), whereas kappa statistics ranged from 0.55 to 1.00 (median 0.87).

Influence of reported study design characteristics: univariable analyses of individual characteristics

The main analyses were based on a data set containing 1973 trials included in 234 meta-analyses. Based on 944 trials from 112 informative meta-analyses, intervention effect estimates were exaggerated by an average of 11% in trials with inadequate or unclear sequence generation

(ROR 0.89, 95% CrI 0.82 to 0.96), and between-trial heterogeneity was higher among such trials ($\kappa = 0.16$, 95% CrI 0.03 to 0.27). When analyses were stratified according to type of outcome measure, the average effect of bias associated with inadequate or unclear sequence generation was greatest for subjective outcomes (ROR 0.83, 95% CrI 0.74 to 0.94) and the increase in between-trial heterogeneity was also greatest for such outcomes ($\kappa = 0.20$, CrI 0.03 to 0.32). There was little evidence that inadequate or unclear sequence generation was associated with exaggeration of intervention effects for all-cause mortality (ROR 0.89, 95% CrI 0.75 to 1.05) or for other objective outcomes (ROR 0.99, 95% CrI 0.84 to 1.16). For all types of outcome, there was only limited between-meta-analysis heterogeneity in mean bias (estimated φ between 0.04 and 0.07).

Based on 1292 trials from 146 informative meta-analyses, intervention effect estimates were exaggerated by 7% in trials with inadequate or unclear allocation concealment (ROR 0.93, 95% CrI 0.87 to 0.99), and between-trial heterogeneity was increased for such studies ($\kappa = 0.12$, 95% CrI 0.02 to 0.23). The effect of inadequate or unclear allocation concealment was greatest among meta-analyses with a subjectively assessed outcome measure (ROR 0.85, 95% CrI 0.75 to 0.95; $\kappa = 0.20$, 95% CrI 0.02 to 0.33). In contrast, the average effect of inadequate or unclear allocation concealment was close to the null for meta-analyses with a mortality outcome (ROR 0.98, 95% CrI 0.88 to 1.10) and other objective outcomes (ROR 0.97, 95% CrI 0.85 to 1.10). Estimates of both between-trial and between-meta-analyses heterogeneity in bias were lower for such outcomes than for subjectively assessed outcomes.

Based on 1057 trials from 104 informative meta-analyses, lack of, or unclear, double blinding was associated with an average 13% exaggeration of intervention effects (ROR 0.87, 95% CrI 0.79 to 0.96). Between-trial heterogeneity was higher in such studies ($\kappa = 0.14$, 95% CrI 0.02 to 0.30), and average bias varied between meta-analyses ($\varphi = 0.14$, 95% CrI 0.03 to 0.28). Average bias (ROR 0.78, 95% CrI 0.65 to 0.92), between-trial heterogeneity ($\kappa = 0.37$, 95% CrI 0.19 to 0.53) and between-meta-analysis heterogeneity in average bias ($\varphi = 0.23$, 95% CrI 0.04 to 0.44) were all greatest for meta-analyses assessing subjective outcome measures. Among meta-analyses with subjectively assessed outcome measures, the effect of lack of blinding appeared greater than the effect of inadequate or unclear sequence generation or allocation concealment.

Influence of reported study design characteristics: univariable analyses of combinations of characteristics

Estimates of the influence of any risk of selection bias (inadequate or unclear sequence generation or allocation concealment, compared with other trials) on intervention effects and heterogeneity were based on 53 informative meta-analyses containing 534 trials, of which 89 (17%) were assessed as at low risk. Risk of selection bias was associated with an average 11% exaggeration of intervention effect estimates (ROR 0.89, 95% CrI 0.78 to 1.00) and with increased between-trial heterogeneity ($\kappa = 0.12$, 95% CrI 0.02 to 0.27). Average effects did not differ substantially according to type of outcome measure.

Only 37 informative meta-analyses [409 trials, of which 58 (14%) were assessed as being at low risk of bias] contributed to the analysis of any risk of bias (inadequate or unclear sequence generation or allocation concealment, or lack of or unclear double blinding, compared with all other trials). Any risk of bias was associated with an average 21% exaggeration of intervention effect estimates (ROR 0.79, 95% CrI 0.64 to 0.92). The numbers of informative meta-analyses and trials included in analyses stratified by type of outcome measure were small, but the influence of any risk of bias appeared the smallest for all-cause mortality outcomes.

A total of 104 informative meta-analyses [990 trials, of which 259 (26%) were assessed as at low risk of bias] contributed to analyses of the influence of inadequate or unclear allocation concealment or lack of double blinding (compared with adequate allocation concealment and

presence of double blinding). Intervention effects from trials at high risk of bias according to this definition were exaggerated by an average 12% (ROR 0.88, 95% CrI 0.81 to 0.95). The ROR was smaller for all-cause mortality outcomes than for other objective or subjective outcome measures. The increase in between-trial heterogeneity appeared greatest for subjective outcome measures ($\kappa = 0.17$, 95% CrI 0.02 to 0.31).

Influence of reported study design characteristics: multivariable analyses

Multivariable analyses of the influence of inadequate or unclear allocation concealment and lack of or unclear double blinding were based on 169 informative meta-analyses (1456 trials) in which both characteristics were assessed. Estimated RORs were similar to, or modestly attenuated compared with, those in the univariable analyses. Estimated influence on heterogeneity was also modestly attenuated.

Estimated RORs from multivariable analyses of the effects of all three characteristics were of similar magnitudes to those in the univariable analyses for each characteristic. For inadequate or unclear sequence generation or allocation concealment, estimated increases in between-trial heterogeneity (quantified by κ) were smaller in multivariable analyses than in the corresponding univariable analyses. Estimates of between-meta-analysis variability in average bias were changed little compared with univariable analyses.

Analyses according to type of intervention and clinical area

For pregnancy and childbirth – the clinical area contributing most meta-analyses to the combined data set – RORs were further from 1 than in the analyses of the whole data set, whereas estimates of the influence of reported study design characteristics on heterogeneity were broadly consistent with analyses of the whole data set. For mental health and circulatory system conditions, RORs were attenuated towards 1. Only small numbers of meta-analyses contributed to estimation of κ, but estimated values of κ and φ were generally smaller for circulatory system meta-analyses than for the other two clinical areas.

The majority of meta-analyses included in the full data set addressed pharmacological interventions; it was therefore unsurprising that overall results restricted to such interventions were consistent with those from the full data set. For surgical interventions, effects of inadequate or unclear sequence generation and inadequate or unclear allocation concealment were estimated from only six and nine meta-analyses respectively; confidence intervals were too wide to allow substantive conclusions to be drawn.

Downweighting potentially biased evidence in future meta-analyses

We investigated the implications of our results for downweighting of potentially biased evidence in future meta-analyses. Because estimated values of κ and φ were greatest for meta-analyses with subjectively assessed outcomes, the minimum variance of the estimated intervention effect for a trial at high or unclear risk of bias is greatest for such trials. Across all Bias in Randomised and Observational studies (BRANDO) trials with inadequate or unclear sequence generation, bias adjustment led to a median 10% [interquartile range (IQR) 4% to 23%] increase in trial-level variance. Downweighting based on results specific to type of outcome measure has the greatest effect in trials with subjectively assessed outcomes [median 20% (IQR 8% to 39%) increase in variance]. Results were broadly similar for downweighting based on inadequate or unclear allocation concealment. The median increase in variance for trials with subjectively measured outcomes that were not double blind or had unclear blinding status was 63% (IQR 22% to 138%).

Downweighting of all trials with inadequate or unclear sequence generation led to a median 13% (IQR 5% to 32%) increase in the variance of the summary (meta-analytic) intervention effect estimate among informative meta-analyses in the BRANDO database. This is in contrast to a

median increase of 217% (IQR 87% to 482%) when such trials are excluded from meta-analyses, because only 26% of trials were assessed to have adequate sequence generation. Bias adjustment for meta-analyses with subjectively assessed outcome measures led to a median 31% (IQR 11% to 56%) increase in the variance of the summary intervention effect estimate, which was again small compared with complete exclusion of such trials. Results were broadly similar for the other study design characteristics, although differences between the effects of downweighting and excluding trials at high or unclear risk of bias were smaller for double blinding, because 56% of trials from informative meta-analyses were double blind.

Conclusions

Bias associated with specific reported study design characteristics of RCTs leads to exaggeration of intervention effect estimates and increases in between-trial heterogeneity. For each of the three characteristics assessed, these effects appeared greatest for subjectively assessed outcome measures. Assessments of the risk of bias in trial results should account for these findings. Downweighting trials at high risk of bias in future meta-analyses, based on these empirical findings, could be an alternative to completely excluding such trials from meta-analyses, resulting in a smaller loss of precision. The strategies used to combine data from several independent data sources and to remove overlap between meta-analyses may be of use for future empirical research.

Recommendations for future research

- Tools for assessing risk of bias in results of RCTs should account for the findings of this study.
- Practical and acceptable methods for correcting and downweighting the results of trials at high risk of bias in new meta-analyses should be developed.
- The influence of further study design characteristics should be explored in new meta-epidemiological studies.
- As far as possible, clinical decisions should not be based on trials in which blinding is not feasible and outcome measures are subjectively assessed.

Funding

Chapter 1

Introduction

Although meta-analyses of randomised trials offer the best evidence for the evaluation of clinical interventions, they are not immune from bias.[1] Bias may arise because of selective reporting of whole trials or of outcomes within trials,[2–5] or because of study design characteristics that compromise internal validity.[6] The design of randomised controlled trials (RCTs) should incorporate characteristics that avoid biases resulting from lack of comparability of the intervention and control groups. For example, concealment of randomised allocation avoids selection bias (differences in the probability of recruitment to the intervention and control groups based on participant characteristics at the time of recruitment), blinding of participants and trial personnel avoids performance bias (differences in aspects of patient management between intervention groups) and blinding of outcome assessors avoids detection bias (differences in outcome measurement).

Empirical evidence about the magnitude and relative importance of the influence of reported study design characteristics on trial results comes from 'meta-epidemiological' studies in which collections of meta-analyses are used to study associations between trial characteristics and intervention effect estimates.[7,8] There is evidence from such studies that inadequate allocation concealment and lack of blinding lead, on average, to exaggeration of intervention effect estimates.[9–11] However, the evidence is not consistent: some studies did not find evidence of effects of these characteristics,[12–14] whereas others suggested that other study design characteristics were of importance.[9–11,15] Possible reasons for this lack of consistency include differential effects of study design characteristics across different clinical interventions or between different objectively and subjectively assessed outcome measures,[16] differences in definitions of characteristics and methods of assessment (some studies relied on assessments reported in contributing systematic reviews[9]), and chance.[8] The extent of overlap between meta-analyses contributing to the different studies is unclear.

As the effects of study design characteristics tend to be estimated imprecisely within individual meta-analyses,[8] large collections of meta-analyses are needed to estimate effects of these characteristics with precision, and to examine variability in these effects according to clinical area or type of outcome measure. It is, therefore, desirable to combine the collections of meta-analyses that were assembled in previous studies so that unified analyses can be conducted. However, these studies assembled their collections of meta-analyses independently, and so the extent of overlap between the meta-analyses and trials included in the different studies is unknown. Identification and resolution of such overlaps is difficult because of the multiplicities inherent in the data structure: trials report results on multiple outcomes and may be included in multiple meta-analyses, while systematic reviews may contain many different meta-analyses. Data are extracted from multiple publications that may describe trials, systematic reviews or both.

By combining data from 10 meta-epidemiological studies into a single database we identified and removed overlaps between trials and meta-analyses, and investigated the consistency of assessments of reported study design characteristics between the different contributing studies.[17] Based on seven contributing studies in which both study design characteristics and results were recorded, we investigated the influence of different study design characteristics on both average intervention effects and between-trial heterogeneity, according to type of intervention and type of outcome measure.

Chapter 2

Development of a combined database for meta-epidemiological research

Methods

Data

Data were combined as part of the BRANDO (Bias in Randomized and Observational studies) project. The authors of 10 meta-epidemiological studies[9,11–15,18–21] (*Table 1*) agreed to contribute their original data to a combined database that would be used to conduct combined analyses and investigate reasons for differences between the results of the original studies. The combined database was created using Microsoft Access™ software (Microsoft Corporation, Redmond, WA, USA). Most authors supplied separate tables containing data on included trials and meta-analyses, allowing linkage between these two levels of data. When such information was not supplied, information from publications of relevant systematic reviews was used to link trials and meta-analyses.[15] Citations (of publications from which the data were extracted) were linked to the data sets using matching via source systematic review, author name and publication year.[12–15] Studies by Sampson *et al.*[20] and McAuley *et al.*[18] and the part of the study by Egger *et al.*[9] that was based on published journal articles [referred to hereafter as Egger (journal)] did not include information on study design characteristics in the included trials, whereas the study by Royle and Milne[19] did not include outcome data. These studies cannot contribute to meta-epidemiological analyses, but were retained to contribute to other, descriptive analyses.

TABLE 1 Characteristics of meta-epidemiological studies that contributed data to the study

Contributing meta-epidemiological study	Number of contributed meta-analyses (trials)	Clinical areas/types of interventions	Number of meta-analyses (trial results) in final database
Als-Nielsen *et al.*[12,22]	48 (523)	Various	46 (506)
Balk *et al.*[13]	26 (276)	Circulatory, paediatrics, infection, surgery	23 (251)
Contopoulos-Ioannidis *et al.*[21]	16 (133)	Mental health	11 (94)
Egger *et al.*[9]	165 (1776)[a]	Various	121 (1115)[a]
Kjaergard *et al.*[15]	14 (190)	Various	8 (72)
McAuley *et al.*[18]	31 (454)	Various	18 (205)[b]
Pildal *et al.*[14]	68 (474)	Various	67 (460)
Royle and Milne[19]	29 (541)	Various	28 (452)[c]
Sampson *et al.*[20]	24 (257)	Circulatory, digestive, mental health, pregnancy and childbirth	14 (112)[b]
Schulz *et al.*[11]	33 (250)	Pregnancy and childbirth	27 (210)

a Data on study design characteristics were available for only 122 contributed meta-analyses (or for 93 meta-analyses in the final database).
b Data on study design characteristics were not available in this study.
c 2×2 data were not available in this study.

Assigning identification numbers and creating a unified data set

To enable automated identification of meta-analyses and trials that occurred in more than one meta-epidemiological study, we labelled each trial and review entry in contributing data sets with a MEDLINE, EMBASE or ISI Web of Science unique publication identifier (ID), in respective order of preference. We first searched MEDLINE (through PubMed™) using bibliographic information for every trial and review included in contributing data sets [e.g. full title, author(s), publication year, journal, volume, pagination] and retrieved their unique MEDLINE identifier (PMID). For references that were not located in MEDLINE, EMBASE was searched (through Ovid™) in the same way and EMBASE-unique identifiers were retrieved. For references that were not located in EMBASE the search was carried out in the ISI Web of Science database, but only a small number of the remaining non-indexed citations were identified in this way. Trial references not indexed in MEDLINE, EMBASE or ISI were cross-checked for duplication with other references using duplicate search facilities in Reference Manager™ bibliographic database software (Thompson Reuters, New York, NY, USA), followed by manual checking. All non-indexed, unique references were assigned a unique identification number derived from their ID number in the Reference Manager database.

We defined in the master database the variables that we wished to combine from the contributing studies. An initial combined data set was then created by mapping each variable in each contributed data set to the predefined variable in the master database that it most closely matched. New variables were then added to the master database to capture contributed data that did not fit the a priori variable definitions.

Identification and removal of duplicate meta-analyses and trials

The unique identifiers of trials included in each meta-analysis in the combined data set were compared with those in every other meta-analysis (regardless of whether or not the meta-analyses assessed the same outcome). Using Intercooled Stata™ version 9 (StataCorp LP, College Station, TX, USA), meta-analyses that contained any trials in common with any other meta-analysis were grouped together. Not all meta-analyses within these sets contained overlapping trials: for example, two meta-analyses with no trials in common might each contain trials in common with a third meta-analysis and these three meta-analyses formed one set.

Meta-analyses that had no overlap with any other meta-analysis in the combined data set were transferred to the final data set. We then considered each set of overlapping meta-analyses in turn. Meta-analyses were excluded from each set until there was minimal overlap between the remaining meta-analyses, according to the following rules:

1. Exclude meta-analyses from the Royle and Milne[19] study, which did not include outcome data (entire meta-analysis removed regardless of the extent of overlap).
2. Exclude meta-analyses for which information on study design characteristics was not available in preference to those for which such information was available. If both meta-analyses in a duplicate pair contain no information on study design characteristics, exclude meta-analyses from earlier studies first: in order Egger (journal),[9] Sampson et al.,[20] McAuley et al.[18] (entire meta-analysis removed regardless of the extent of overlap).
3. Exclude meta-analyses from the portion of the study by Egger et al.[9] that was based on assessment of study design characteristics in Cochrane reviews in preference to meta-analyses from studies that directly assessed these characteristics.
4. Exclude meta-analyses with fewer assessed study design characteristics in preference to those with more assessed characteristics.
5. Exclude meta-analyses from less recently published systematic reviews in preference to more recently published reviews.
6. Exclude meta-analyses including fewer trials in preference to meta-analyses including more trials.

These rules were used in order of priority, that is, the next rule was applied only if the previous could not yield a decision. We recorded reasons for all decisions to exclude meta-analyses. Pairs of meta-analyses with recorded study design characteristics and results that had only a minimal overlap between them were retained at this stage. Overlap was considered minimal if the number of overlapping trials was no more than 10% of the sum of the numbers of overlapping trials and unique trials from both meta-analyses. We then removed the overlapping trials from the meta-analysis in the duplicate pair that (1) contained more trials or (2) according to the rules 3 to 5 above. If these rules did not yield a clear decision, overlapping trials were removed from one of the meta-analyses at random. At both stages, the choice of meta-analyses or trials to be removed was independent of the assessment of the study design characteristics, and disagreements in these assessments between studies were not considered. Similarly, the decision on removal was also independent of the types of outcome measure assessed in the overlapping meta-analyses.

Some meta-analyses contained multiple results from the same trial, usually because they included multiarm trials in which the same control/comparison group was compared with two different treatment groups. Where appropriate, the two treatment groups were combined. In other cases one of the results was removed at random. We then checked the 2×2 results tables from trials in the deduplicated data set against the data in the source review publications. Inconsistencies were clarified with the contributors and corrected where necessary. We recoded the direction of outcome events where necessary so that the coded outcome for each trial corresponded to an adverse (undesirable) event.

Assessing consistency of assessment of reported study design characteristics in different contributing studies

We used trials included in more than one meta-epidemiological study to assess the reliability of the assessment of methodological characteristics between studies. For this analysis we defined such trials as those with the same bibliographic reference occurring in more than one meta-epidemiological study, irrespective of numbers of participants or events, or the type of outcome measure. To assess the inter-rater reliability we compared contributors' assessments of the following three methodological characteristics: adequacy of the method for generating the random sequence used for allocating participants to treatment groups (sequence generation), adequacy of concealment of treatment allocation from participants and investigators at the point of enrolment into the trial (allocation concealment), and contributors' assessments of whether or not a trial was double blind or not (blinding). Kappa statistics were calculated for the assessment of sequence generation (inadequate/unclear compared with adequate), allocation concealment (inadequate/unclear compared with adequate) and blinding (not double blind/unclear compared with double blind) of duplicated trials for each pair-wise comparison between contributing meta-epidemiological studies. Only comparisons with at least 10 overlapping trials between any two meta-epidemiological studies were analysed, for each of the three characteristics.

Classification of interventions and outcome measures

We classified the type of experimental intervention, type of comparison intervention and type of outcome measure for each meta-analysis in the final data set. Interventions were classified based on an expanded and modified version of the classification by Moja et al.[23] Comparison interventions were further classified as inactive (e.g. placebo, standard care) or active. When it was not clear which intervention group should be considered experimental and which the comparison, at least two study collaborators made a consensus decision. In cases when a decision could not be reached, such meta-analyses were excluded from analyses of associations between study design characteristics and intervention effect estimates.

We classified outcome measures according to an expanded and modified version of the classification developed by Wood et al.[16] Outcome measures were further grouped as all-cause mortality, other objectively assessed (including pregnancy outcomes and laboratory outcomes),

objectively measured but potentially influenced by clinician/patient judgement (e.g. hospital admissions, total dropouts/withdrawals, caesarean section, operative/assisted delivery) and subjectively assessed. When a review reported that different methods of outcome assessment were used in different trials within the same meta-analysis, the meta-analysis was classified according to the most subjective method of outcome assessment. This was the case in 16 meta-analyses. For example, some trials of smoking cessation assessed outcomes using exhaled carbon dioxide or salivary cotinine (classified as objectively assessed), whereas for others it was by patient self-report (classified as subjectively assessed). Classifications of interventions and outcome measures were checked by at least two of the collaborators who were clinicians by training (CG, LLG, BA-N or JP). Classifications of both outcome measures and interventions were based solely on the information provided in the review from which the meta-analysis was extracted. We did not retrieve further information on individual outcome measures from publications of included trials.

Results

Database structure

The database design allowed for multiple results from the same trial to be contained in different meta-analyses, multiple meta-analyses in systematic reviews, overlapping meta-analyses between systematic reviews and multiple references to the same trial or review. The final database structure consisted of six tables (*Figure 1*). The first five tables, *References*, *Trials*, *Trial Results*,

References
of trials and systematic review articles:
- authors
- title
- journal
- volume
- year of publication
- PubMed or other ID, etc.

Systematic reviews
Basic information about reviews:
- title/topic
- clinical area according to ICD-10

Trials
Trial identification and methodological characteristics:
- random sequence generation
- allocation concealment
- blinding of participants and personnel
- funding source (subset only)

Relationships
Identifiers for every entry in all of the five tables and linking information between them

Meta-analyses
Characteristics and results for each individual meta-analysis:
- participants
- intervention
- comparison intervention
- outcome measure
- outcome category (four groups)
- pooled summary estimates and CI (as reported in review)
- meta-analysis method used

Trial results/outcomes
- outcome specific, can hold more than one outcome from same trial
- events in intervention group
- events in comparison group
- total in intervention group
- total in comparison group
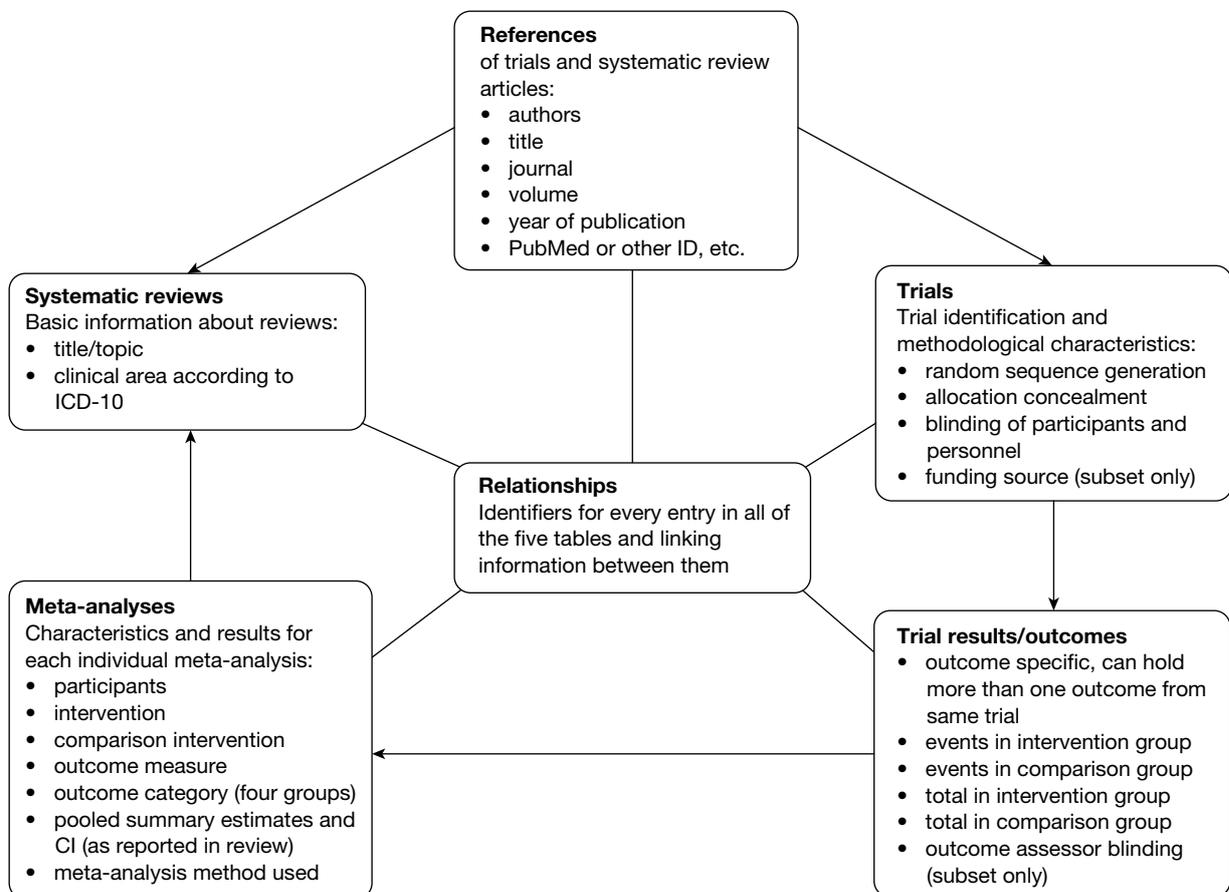- outcome assessor blinding (subset only)

FIGURE 1 Schematic representation of the final database structure. ICD-10, *International Classification of Disease*, 10th edition.[30]

Meta-analyses and *Systematic Reviews*, contained relevant study characteristics. The sixth table, *Relationships*, contained the identifiers necessary to link information between all the other tables. Each reference entry was linked to its related trial or systematic review, and further links were established between trials, trial results, meta-analyses and systematic reviews in the corresponding database tables. *Figure 1* lists core variables included in each table.

Removing overlaps

Figure 2 depicts the derivation of the final database through removal of overlaps between meta-analyses. The initial combined data set contained 427 systematic reviews with 454 meta-analyses, 4857 trials and 4874 trial results. Of the 454 meta-analyses, 196 contained at least one trial that overlapped with at least one other meta-analysis, among which we identified 71 sets of meta-analyses containing overlapping trials. The size of these sets varied from 2 to 17 meta-analyses (median 2). We removed 91 entire meta-analyses containing 1354 trial results during the deduplication process. Of the 1354 results removed, 844 were duplicates of trials retained in the database, whereas 510 were unique. Of those that were unique, 340 (67%) were from studies that did not record either methodological characteristics or results, thus 170 potentially informative trials were removed during this process. A further 24 individual trial results were removed from 16 additional meta-analyses for which overlap was minimal (see methods). One trial was removed from 10 meta-analyses, two trials from four meta-analyses, and three trials from two meta-analyses each. An additional 19 trial results were removed as a result of overlaps within nine of the meta-analyses (e.g. where there were two comparisons with a common control group). The final database contained 352 systematic reviews contributing 363 meta-analyses, 3474 trials and 3477 trial results.

Initial combined data set:
- 427 systematic reviews
- 454 meta-analyses
- 4857 trials
- 4874 trial results

Of 454 meta-analyses in the initial database:
- 258 meta-analyses were unique
- 196 have at least one overlapping trial with at least one other meta-analysis

Deduplication process:
- 71 sets of overlapping meta-analyses identified and assessed

Removed:
- 75 entire systematic reviews
- 91 entire meta-analyses, containing 1354 trial results
- A further 24 trial results were removed from additional 16 meta-analyses where overlap was too small to justify removal of the entire meta-analysis

Deduplicated database:
- 352 systematic reviews
- 363 meta-analyses
Some trials still duplicated **within** meta-analyses (multiarm trials)

Additional 19 trial results removed owing to overlaps within 9 individual meta-analyses

Final database:
- 352 systematic reviews
- 363 meta-analyses
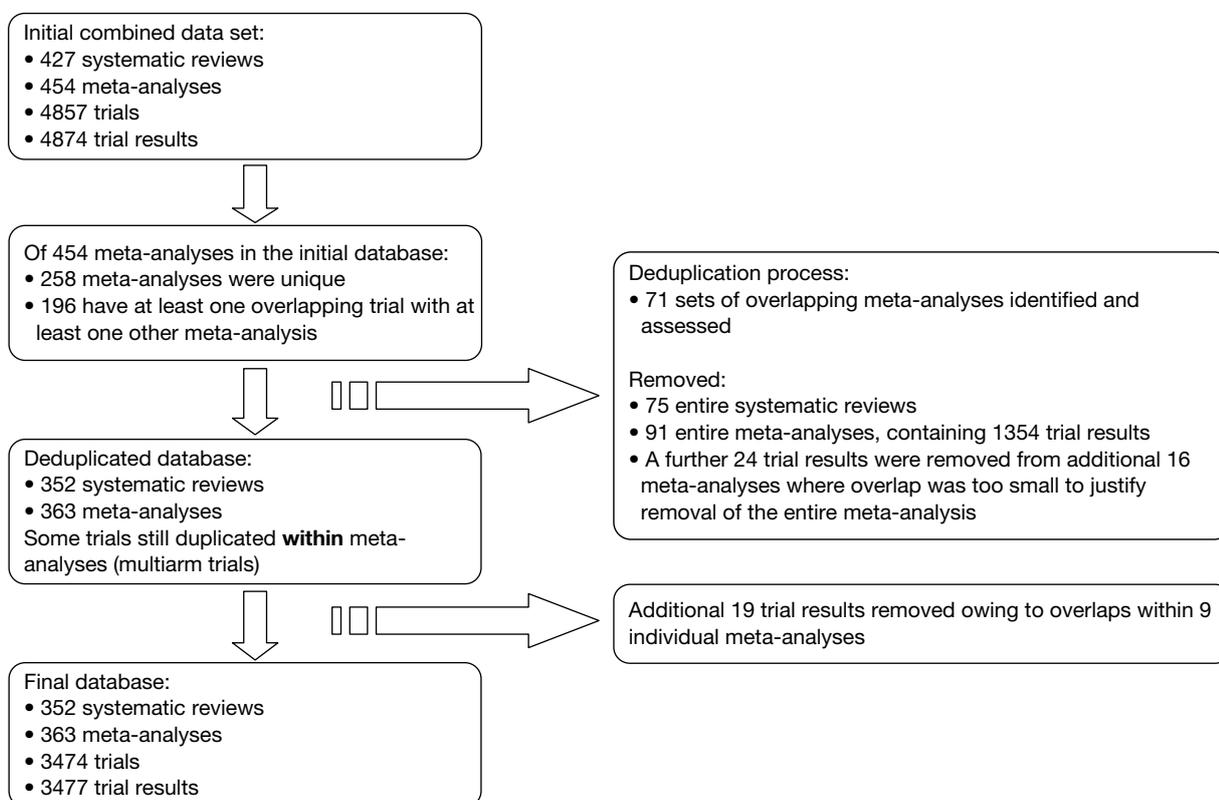- 3474 trials
- 3477 trial results

FIGURE 2 Flow diagram depicting the deduplication process.

Table 2 shows an example of a set of four meta-analyses containing overlapping trials. Three meta-analyses (1005, 2456 and 2531) were taken from the same Cochrane review,[24] whereas 2097 was from a journal review[25] with a similar topic. *Table 3* summarises the overlap between the trials in each pair of meta-analyses. To remove overlaps, meta-analyses were dropped in the following order: (1) meta-analysis 2097, because it was contributed by the Egger (journal) study,[9] which did not include study design characteristics; (2) meta-analysis 2456, because study design characteristics in this study were extracted from a Cochrane review rather than from primary publications. The overlap between the remaining two meta-analyses (1005, from the study by Contopoulos-Ioannidis *et al.*[21] and 2531 from the study by Kjaergard *et al.*[15]) was then examined in detail (*Table 4*). Each meta-analysis contained 10 trials, of which eight overlapped. Of the overlapping trials, seven had the same totals per group (different event numbers), and one trial had different data for the control group. These meta-analyses assessed two different outcomes:

TABLE 2 Example of a set of meta-analyses that contained overlapping trials

Meta-analysis ID	PMID of review	Title of systematic review	Assessed outcome	Meta-epidemiological study
1005	10796543	Risperidone versus typical antipsychotic medication for schizophrenia (Cochrane review–CD000440)[24]	No change in positive and negative syndrome scale	Contopoulos-Ioannidis *et al.*[21]
2097	9097896	Risperidone in the treatment of schizophrenia: a meta-analysis of randomized controlled trials (journal review)[25]	Clinical improvement	Egger (journal)[9]
2456	10796543	Risperidone versus typical antipsychotic medication for schizophrenia (Cochrane review–CD000440)[24]	Withdrawals/dropouts	Egger (CDSR)[9]
2531	10796543	Risperidone versus typical antipsychotic medication for schizophrenia (Cochrane review–CD000440)[24]	Withdrawals/dropouts	Kjaergard *et al.*[15]

CDSR, Cochrane Database of Systematic Reviews.

TABLE 3 Examples of trial overlap between the above meta-analyses – showing different degrees of overlap and decision process

Meta-analysis 1			Meta-analysis 2			Overlaps					
ID	Contributing study	Unique trials	ID	Contributing study	Unique trials	Exact	Same totals	Different	Within	Comment
1005	Contopoulos-Ioannidis *et al.*[21]	3	2097	Egger (journal)[9]	4	3	2	2	0	2097 removed, no methodology data
1005	Contopoulos-Ioannidis *et al.*[21]	3	2456	Egger (CDSR)[9]	5	0	7	0	0	2456 removed, data from review only
1005	Contopoulos-Ioannidis *et al.*[21]	2	2531	Kjaergard *et al.*[15]	2	0	7	1	0	Detailed assessment needed (see *Table 4*)
2097	Egger (journal)[9]	4	2456	Egger (CDSR)[9]	5	0	5	2	0	Both already removed
2097	Egger (journal)[9]	3	2531	Kjaergard *et al.*[15]	2	0	6	2	0	2097 already removed
2456	Egger (CDSR)[9]	3	2531	Kjaergard *et al.*[15]	1	8	0	1	0	2456 already removed

CDSR, Cochrane Database of Systematic Reviews.
Unique trials, number of unique (non-overlapping) trials in meta-analysis. *Exact*, number of trials for which all 2×2 data were the same in the overlapping meta-analyses. *Same totals*, number of trials in which the event data differ, but there was the same total number of participants in the intervention and comparison groups. *Different*, number of trials in which both group total and event data are different in the overlapping meta-analyses. *Within*, number of trials that are duplicated within the individual meta-analysis (usually this was in case of multiarm trials entered as multiple comparisons).

TABLE 4 Pair-wise comparison for meta-analyses 1005 and 2531

Trial ID[a]	Meta-analysis 1005 (Contopoulos-Ioannidis[21]) – kept						Meta-analysis 2531 (Kjaergard[15]) – removed						Trial unique or overlaps?
	d_1	h_1	n_1	d_0	h_0	n_0	d_1	h_1	n_1	d_0	h_0	n_0	
R_17134	65	28	93	77	14	91							Unique
R_17135	126	223	349	170	156	326							Unique
P_1375801	15	7	22	17	5	22	1	21	22	5	17	22	Overlap (equal totals)
P_1381102							26	27	53	31	22	53	Unique
P_7508675	18	37	55	22	30	52	14	41	55	15	37	52	Overlap (equal totals)
P_7514366	139	117	256	46	20	66	122	134	256	38	28	66	Overlap (equal totals)
P_7542829	20	28	48	29	21	50	17	31	48	23	27	50	Overlap (equal totals)
P_7545060	457	679	1136	97	129	226	280	856	1136	63	163	226	Overlap (equal totals)
P_7683702	44	48	92	11	10	21	36	56	92	13	8	21	Overlap (equal totals)
P_7691017	6	10	16	5	14	19	3	13	16	0	19	19	Overlap (equal totals)
P_7694306							0	31	31	3	28	31	Unique
P_8834417	4	17	21	8	12	20	4	17	21	6[b]	4[b]	10[b]	Overlap (different)[b]

a   The origin of the trial IDs is indicated by the prefix, with P_ denoting PMID, E_ denoting EMBASE ID (not seen in this example) and R_ denoting trials not indexed in any of the major databases.

b   Possible data entry error (errors in 2×2 data were not checked or corrected for meta-analyses that were removed in the deduplication process).

Decision: meta-analysis 1005 was retained because the contributing study provided information on one additional study design characteristic (sequence generation).

Event 2×2 data: d, unfavourable events; h, favourable events; *n*, total number of participants per group; subscript '1' denotes experimental intervention group, subscript '0' denotes comparison intervention group.

that in meta-analysis 1005 was 'No change in positive and negative syndrome scale (data greater than 20%)', whereas that in meta-analysis 2531 was 'dropouts'. This explains why there was no exact correspondence in the 2×2 data for any trial. Meta-analysis 1005 was retained because the contributing study provided information on one additional study design characteristic (sequence generation).

Reliability of assessment of reported study design characteristics

Eight of the participating meta-epidemiological studies contained sufficient trials (at least 10) in common to contribute to analyses of the reliability of assessment of reported study design characteristics (*Table 5*). Overall, there was good agreement between assessments carried out in the different contributing studies. For sequence generation (two comparisons), the percentages of studies in which the assessments were in agreement were 81% and 82% and kappa statistics were 0.56 and 0.64. For allocation concealment (12 comparisons), percentage agreement varied between 52% and 100% and kappa statistics between 0.19 and 1.00 (median 0.58). The lowest kappa (0.19) was observed in the comparison between the Egger *et al.*[9] and Pildal *et al.*[14] studies: 10 (48%) studies assessed as having adequate concealment of allocation in the Egger study (which used assessments reported by Cochrane review authors) were assessed as having inadequate or unclear concealment of allocation in the Pildal study. Assessments were most reliable for blinding (nine comparisons): percentage agreement ranged from 80% to 100% (in four comparisons), and kappa statistics ranged from 0.55 to 1.00 (median 0.87).

Interventions and outcome measures

Outcome measures were initially classified into one of 23 categories. *Table 6* shows the numbers of meta-analyses with outcomes in each of these categories, together with the final category grouping. The most common outcome measure was all-cause mortality (64 meta-analyses, 18%), followed by clinician-assessed outcomes (51 meta-analyses, 14%).

TABLE 5 Assessor agreement and kappa statistics for studies with at least 10 trials per comparison

Study 1	Study 2	Kappa	Agreement (%)	No. of trials	$n00$[a]	$n01$[a]	$n10$[a]	$n11$[a]
Sequence generation (inadequate/unclear vs adequate)								
Als-Nielsen *et al.*[12,22]	Pildal *et al.*[14]	0.64	82	22	11	0	4	7
	Schulz *et al.*[11]	0.56	81	16	10	0	3	3
Allocation concealment (inadequate/unclear vs adequate)								
Als-Nielsen *et al.*[12,22]	Balk *et al.*[13]	0.63	93	14	12	1	0	1
	Egger *et al.*[9]	0.78	96	94	82	2	2	8
	Kjaergard *et al.*[15]	0.48	86	56	43	6	2	5
	Pildal *et al.*[14]	0.58	86	22	16	2	1	3
	Royle and Milne[19]	0.63	85	54	36	8	0	10
	Schulz *et al.*[11]	0.76	94	16	13	1	0	2
Egger *et al.*[9]	Contopoulos-Ioannidis *et al.*[21]	NA	100	14	14	0	0	0
	Pildal *et al.*[14]	0.19	52	21	8	0	10	3
Kjaergard *et al.*[15]	Balk *et al.*[13]	0.39	67	15	6	0	5	4
	Egger *et al.*[9]	0.58	90	63	52	0	6	5
	Contopoulos-Ioannidis *et al.*[21]	0.38	80	10	7	1	1	1
Schulz *et al.*[11]	Egger *et al.*[9]	1.00	100	20	14	0	0	6
Blinding (not double blind/unclear vs double blind)								
Als-Nielsen *et al.*[12,22]	Egger *et al.*[9]	0.89	95	112	63	4	2	43
	Kjaergard *et al.*[15]	0.74	88	56	20	5	2	29
	Pildal *et al.*[14]	1.00	100	22	10	0	0	12
	Schulz *et al.*[11]	NA	100	15	15	0	0	0
Egger *et al.*[9]	Contopoulos-Ioannidis *et al.*[21]	1.00	100	17	10	0	0	7
	Pildal *et al.*[14]	0.87	95	19	5	0	1	13
Kjaergard *et al.*[15]	Egger *et al.*[9]	0.69	84	77	32	6	6	33
	Contopoulos-Ioannidis *et al.*[21]	0.55	80	10	2	2	0	6
Schulz *et al.*[11]	Egger *et al.*[9]	NA	100	15	15	0	0	0

NA, not applicable (not possible to calculate kappa as all trials were assessed as 'adequate' or 'not double blind' in both studies).

a　$n00$, number of trials judged 'inadequate/unclear' or 'not double blind' in both contributing studies; $n01$, number of trials judged 'inadequate/unclear' or 'not double blind' in study 1 and 'adequate' or 'double blind' in study 2; $n10$, number of trials judged 'adequate' or 'double blind' in study 1 and 'inadequate/unclear' or 'not double blind' in study 2; $n11$, number of trials judged 'adequate' or 'double blind' in both contributing studies.

Experimental and comparison interventions were classified as shown in *Table 7*. The majority of included meta-analyses (242 meta-analyses, 67%) assessed a pharmacological experimental intervention. The most common comparison intervention was placebo or no treatment (225 meta-analyses, 62%), followed by a pharmacological comparison (54 meta-analyses, 15%). Forty-eight meta-analyses (397 trial outcomes) in which it was unclear what was the experimental intervention were excluded from meta-epidemiological analyses.

Description of the final database

The final data set contained information from 363 meta-analyses containing 3474 trials. Of these, 282 meta-analyses (2572 trials) had 2×2 results data available. A total of 186 meta-analyses (1236 trials) had information on randomisation sequence generation, 228 (1840) on allocation concealment and 234 (1970) on blinding. In total, 175 meta-analyses (1171 trials) had information on 2×2 results data and these three study design characteristics.

TABLE 6 Classification of meta-analysis outcome measures

Type of outcome measure	No. of meta-analyses	Outcome group[a]
Adverse events (as adverse effects of the treatment)	6	4
All-cause mortality	64	1
Cause-specific mortality	2	4
Clinician-assessed outcomes (e.g. body mass index, blood pressure, lung function, infant weight)	51	Mostly 4
Composite end point including end points other than mortality/major morbidity	0	NA
Composite end point including mortality and/or major morbidity	9	2, 3 or 4
Global improvement	4	4
Health perceptions (person's own view of general health)	0	4
Laboratory-reported outcomes (e.g. blood components, tissue analysis, urinalysis)	29	Mostly 2 (two 4)
Lifestyle outcomes (including diet, exercise, smoking)	12	Mostly 4 (one 2)
Major morbidity event (including myocardial infarction, stroke, haemorrhage)	6	4
Mental health outcomes (including cognitive function, depression and anxiety scores)	16	4
Other outcomes (not classified elsewhere)	7	2, 3 or 4
Pain (extent of pain a patient is experiencing)	13	4
Perinatal outcomes	32	2 or 3
Pregnancy outcomes	11	2
Quality of life (including ability to perform physical, daily and social activities)	0	4
Radiological outcomes (including radiograph abnormalities, ultrasound, magnetic resonance imaging results)	12	4
Resource use (including cost, hospital stay duration, number of procedures)	4	3
Satisfaction with care (including patient views and clinician assessments)	0	4
Surgical and device-related outcomes	16	Mostly 4 (two 3)
Symptoms or signs of illness or condition	35	4
Withdrawals/dropouts/compliance	16	3

NA, not applicable.

a   1, all-cause mortality; 2, other objectively assessed; 3, objectively measured, but potentially influenced by clinician/patient judgement; 4, subjectively assessed. Meta-analyses that included trials with both objectively and subjectively assessed outcomes were classified as having subjectively assessed outcomes. Outcome information was not available for 18 meta-analyses from Royle and Milne.[19]

TABLE 7 Categories of experimental and comparison interventions

Intervention categories	Meta-analyses per category[a]	
	Experimental	Comparison
Experimental and comparison interventions		
Diagnostics and screening	7	4
Interventions applying energy source for therapeutic purposes (e.g. ECT/radiotherapy, light therapy, etc.)	5	1
Lifestyle interventions (diet change, exercise, smoking cessation, etc.)	5	1
Medical devices	15	11
Pharmacological	242	54
Physical and manipulative therapy (physiotherapy, chiropractics, etc.)	2	0
Psychosocial (including psychotherapy, counselling, behavioural, advice, guidelines, self-help, etc.)	16	2
Resources and infrastructure/provision of care	9	0
Specialist nutritional interventions and fluid delivery (e.g. parenteral nutrition)	8	6
Surgical interventions or procedures	26	19
Therapies of biological origin (excluding vaccines, including in vitro fertilisation)	2	1
Vaccines	4	1
Other	7	3
Dual intervention (a complex intervention with components that fall in two different categories)	12	0
Multiple intervention (a complex intervention with components that fall in more than two different categories)	3	2
Comparison interventions only		
No treatment		58
Placebo		99
Placebo or no treatment combined at meta-analysis level		68
Standard/usual care		11
Standard care and/or placebo and/or no treatment combined at meta-analysis level		8

ECT, electroconvulsive therapy.
a   Information on comparison intervention was not available for 14 meta-analyses from Royle and Milne.[19]

Chapter 3

Influence of reported study design characteristics on average intervention effects and between-trial heterogeneity

Introduction

In this chapter we report the results of analyses of the influence of three reported study design characteristics [inadequate or unclear (compared with adequate) random sequence generation; inadequate or unclear (compared with adequate) allocation concealment; and absent or unclear double blinding (compared with double blinding)] on both average intervention effects and between-trial heterogeneity, according to the type of intervention and type of outcome. We examine whether or not these influences vary with the type of clinical area, intervention, comparison and outcome measure, examine effects of combinations of study design characteristics, estimate-adjusted effects using multivariable models, compare results with those derived using previously used (meta-meta-analytic) methods and explore implications of these findings for downweighting of trials whose study design characteristics are associated with bias in future meta-analyses.

Methods

For this part of the study we removed, from the database described in *Chapter 2*, data from three meta-epidemiological studies[18–20] and one part of the Egger *et al.* study[9] that did not collect data on study design characteristics (87 meta-analyses with 1093 trials). We also removed 36 meta-analyses (300 trials) in which it was not possible to classify one intervention as experimental and the other as control, one meta-analysis (four trials) that had a continuous outcome, 45 trials in which outcome data were missing and 50 trials in which either no or all participants experienced the outcome event.

Categories of intervention (see *Table 7*) containing fewer than 10 meta-analyses or 50 trials were combined into four types of intervention: pharmacological, surgical, psychosocial and behavioural, and all other interventions. Comparison interventions were classified as inactive (e.g. placebo, no intervention, standard care) or active. Outcome measures were grouped as all-cause mortality, other objectively assessed (including pregnancy outcomes and laboratory outcomes), objectively measured but potentially influenced by clinician/patient judgement (e.g. hospital admissions, total dropouts/withdrawals, caesarean section, operative/assisted delivery) and subjectively assessed (e.g. clinician-assessed outcomes, symptoms and symptom scores, pain, metal health outcomes, cause-specific mortality). Too few meta-analyses were categorised as having 'other objectively assessed' and 'objectively measured but potentially influenced by clinician/patient judgement' outcome measures to allow separate analyses; these categories were therefore grouped together.

Statistical methods

Intervention effects were modelled as log-odds ratios and outcomes were recoded where necessary so that odds ratios (ORs) < 1 corresponded to beneficial intervention effects. We fitted Bayesian hierarchical bias models using the formulation previously described as 'Model 3' by Welton *et al.*[26] We assumed that the observed number of events in each arm of each trial has a binomial distribution, with the underlying log-odds ratio in trial i in meta-analysis m ($LORim$) equal to

$$LOR_{im} = \delta_{im} + \beta_{im}X_{im} \qquad \text{[Equation 1]}$$

where $X_{im} = 1$ and 0 for trials with and without the reported characteristic. The parameter δ_{im} represents the intervention effect in trial i of meta-analysis m. These are assumed to be randomly distributed with variance τ^2_m within each meta-analysis:

$$\delta_{im} \sim Normal\left(d_m, \tau^2_m\right) \qquad \text{[Equation 2]}$$

Parameter β_{im} quantifies the potential bias associated with the study design characteristic of interest. We assumed the following model structure:

$$\beta_{im} \sim Normal\left(b_m, \kappa^2\right) \qquad \text{[Equation 3]}$$

for trials i with the reported characteristic in each meta-analysis m, and

$$b_m \sim Normal\left(b_0, \varphi^2\right) \qquad \text{[Equation 4]}$$

across meta-analyses.

This allows for three effects of bias. First, mean intervention effects may differ between trials with and without the reported study design characteristic. Estimated mean differences (b_0) were exponentiated and are thus reported as ratios of odds ratios (RORs). Second, variation in bias between trials within meta-analyses is quantified by standard deviation κ; κ^2 corresponds to the average increase in between-trial heterogeneity in trials with a specified study design characteristic. Third, variation in mean bias between meta-analyses is quantified by between-meta-analysis standard deviation φ. We derived 95% credible intervals (CrI) for each parameter. In presenting results from our primary analyses, we also display the posterior variance of the parameter b_0, denoted by V_0. Use of this value in downweighting results from trials at high risk of bias in future meta-analyses is discussed below. For the results of our secondary analyses we do not present V_0 explicitly, but the posterior uncertainty about b_0 is reflected in the CrI for the ROR.

Data management and cleaning prior to analysis, and graphical displays of results, were carried out using Stata Version 11. Bias models were then fitted using WinBUGS Version 1.4 (MRC Biostatistics Unit, Cambridge, UK),[27] assuming vague prior distributions for unknown parameters. Preliminary results indicated that estimated variance components κ and φ were sensitive to the prior distributions assumed for these parameters. Sensitivity to priors for variance parameters is a well-known problem in Bayesian hierarchical modelling.[28,29] This motivated a simulation study in which the performance of a range of prior distributions for variance components was compared, assuming typical values from the BRANDO database (Harris *et al.*, Health Protection Agency, 2010, personal communication). The prior found to give the best overall performance was a modified Inverse Gamma(0.001, 0.001) prior with increased weight on small values. This prior was therefore assumed for each variance parameter in all analyses.

For location parameters (overall mean bias, baseline response rates, treatment effects), Normal(0, 1000) priors were assumed.

Meta-analyses can inform estimates of the effect of a study design characteristic only if they contain at least one trial with and one without the characteristic. We refer to such meta-analyses and trials from these meta-analyses as informative. As it is impossible to estimate both τ^2 and κ in a meta-analysis with fewer than two studies with and without the study design characteristic of interest, such meta-analyses were prevented from contributing to the estimation of κ by use of the 'cut' function in WinBUGS.

We first conducted univariable analyses for each study design characteristic separately using all informative meta-analyses for that characteristic. The primary analysis used dichotomised variables for each characteristic (inadequate/unclear compared with adequate for sequence generation and allocation concealment, and not double blind/unclear compared with double bind). All such analyses were repeated separately for different types of outcome measure (all-cause mortality, other objectively assessed and subjectively assessed). Evidence that effects of bias differed according to the type of outcome was quantified using posterior probabilities that effects for subjective or other objective outcomes were larger than those for mortality outcomes: for example $\Pr(\kappa_{subjective} > \kappa_{mortality})$. The main univariable analyses were repeated using the meta-meta-analytic approach used in previous analyses[8,16] allowing for random effects both within and between meta-analyses.

Further univariable and multivariable analyses were conducted using two data subsets: (1) meta-analyses of trials with information on all three study design characteristics and (2) meta-analyses of trials with information on both allocation concealment and blinding. Subset 2 was used because many studies did not have a recorded bias judgement on sequence generation (see *Table 11*). We conducted univariable analyses on three composite dichotomous variables: risk of bias due to inadequate/unclear allocation concealment or lack of blinding (using subset 2), any risk of selection bias (inadequate or unclear sequence generation or allocation concealment) and any risk of bias (inadequate or unclear sequence generation or allocation concealment, or not double blind). Analyses of the second two composite variables used subset 1.

Multivariable analyses were based on an extension of 'Model 3' of Welton *et al.*[26] We assumed distinct variance components associated with each study design characteristic. In the main multivariable analyses, we assumed no interactions between the different characteristics. In an additional analysis on subset 2 we allowed for interactions between inadequate allocation concealment and lack of double blinding. Interaction terms were assumed to have the same hierarchical structure as the main effects, again with distinct variance components. The implied average bias in studies with both characteristics was estimated (on the log-odds scale) as the sum of the fitted coefficients representing the average effect of each of the two characteristics and the fitted coefficient representing the average interaction term. A 95% CrI for this sum, accounting for correlations between the three coefficients, was calculated using WinBUGS. These measures were exponentiated in order to express the implied average bias as ROR. For comparison, we also calculated the corresponding implied average bias for the model without interaction terms for subset 2. We repeated all univariable analyses on subsets 1 and 2 to allow comparisons with the results of multivariable analyses.

In additional analyses we estimated separate effects of 'inadequate' and 'unclear', for each study design characteristic, by fitting models in which these two categories had different average bias (compared with 'adequate' trials) and distinct variance components (κ and φ). It was necessary to exclude some of the contributing meta-epidemiological studies from these analyses because their original data coding had not separated unclear from inadequate (see *Tables 9* and *11*). We

conducted separate analyses according to clinical area and type of intervention, and repeated analyses in meta-analyses that were derived from the subset of contributing studies not included in the study of Wood *et al.*[16] For meta-analyses comparing two active interventions it is not possible to estimate RORs quantifying average bias, or between-meta-analysis heterogeneity in average bias, because there is no clear direction in which bias operates. Instead, we estimated parameters from restricted models to estimate increases in between-trial (within-meta-analysis) heterogeneity in trials with, compared with those without, specified study design characteristics, among meta-analyses containing at least two trials with and without the characteristic of interest.

Welton *et al.*[26] showed how results from hierarchical bias models can be used to formulate a prior distribution for the bias associated with a study design characteristic in a new fixed-effect meta-analysis that is assumed to be statistically exchangeable with the meta-analyses used to estimate the model parameters. This approach was based on a normal approximation to the distribution of the observed intervention effect y_i in each new trial i, which is assumed to have known sampling variance σ_i^2. Assuming known b_0, κ, φ and V_0, a posteriori use of the empirically based prior distribution leads to results from trials with the reported characteristic being corrected for the estimated average bias across meta-analyses (estimated b_0), and the variance of such results being increased from σ_i^2 to $(\sigma_i^2 + \kappa^2 + \varphi^2 + V_0)$. The minimum variance of the estimated intervention effect that can in theory be achieved by an infinitely large trial with the reported characteristic is therefore $\kappa^2 + \varphi^2 + V_0$.

It is of interest to quantify the likely magnitude of increases in variance resulting from application of this bias adjustment to future trial results. To do so, for each trial in the BRANDO data set we calculated the observed log-odds ratio y_i and the Woolf estimate of its sampling variance, σ_i^2. We assumed that these represented a typical range of σ_i^2s that may be observed in future trials. For each high/unclear-risk trial result in turn, we calculated the percentage increase in variance that would result from bias adjustment at the trial level, that is:

$$100 \times \frac{\sigma_i^2 + \kappa^2 + \varphi^2 + V_0}{\sigma_i^2} \% \qquad \text{[Equation 5]}$$

The calculations used the posterior median values of κ and φ. We summarised these percentage increases in variance by the median and interquartile range (IQR) across trials for each study design characteristic.

Formulae from Welton *et al.*[26] also allow us to calculate a bias-adjusted summary mean and variance of the intervention effect in a new fixed-effect meta-analysis. Using this approach, results from trials at low risk of bias are assigned the usual inverse variance ($1/\sigma_i^2$) weight. For trials at high/unclear risk of bias, the bias-corrected estimated effect size is used, and downweighted according to a function of σ_i^2, κ^2, φ^2 and V_0 (interested readers are referred to page 123 of Welton *et al.*[26]). The resulting bias-adjusted estimate of the summary effect size will therefore have a larger variance than the standard unadjusted meta-analytic summary. The magnitude of this increase in variance will depend on the number of trials and the variances of the intervention effect estimates in the new meta-analysis, and the number of trials classified as high/unclear risk of bias. We assumed that the meta-analyses in the BRANDO database are typical in terms of these characteristics, and calculated the percentage increase in variance of the summary log-odds ratio due to bias adjustment for each meta-analysis in turn. These were summarised by the median and IQR across meta-analyses for each study design characteristic.

Results

Table 8 shows the included meta-epidemiological studies, the sources of their collections of meta-analyses and the study design characteristics that they assessed. For five studies[12–15,21] data from each trial report were extracted by two researchers independently; in the study by Pildal *et al.*[14] the assessors were also blinded to trial results. In the study by Schulz *et al.*,[11] one researcher, who was blinded to the trial outcome, assessed the reported methodological characteristics of included trials using a detailed classification scheme. The study of Egger *et al.*[9] was based on quality assessments by the authors of the included Cochrane reviews, which were generally carried out in duplicate by two observers.

Table 9 shows the definitions of adequate sequence generation, allocation concealment and blinding used in the seven studies. Definitions of adequate sequence generation and adequate allocation concealment were similar in all seven studies and were based on the definitions originally proposed by Schulz *et al.*[11] Sequence generation was assessed as adequate, unclear or inadequate in five studies. The study by Kjaergard *et al.*[15] provided only dichotomised assessments of adequate compared with inadequate or unclear sequence generation, and Balk *et al.*[13] did not assess adequacy of sequence generation. Allocation concealment was assessed as adequate, unclear or inadequate in all seven studies. Definitions of 'double blind' varied between studies and were somewhat stricter in the studies by Schulz *et al.*[11] and Pildal *et al.*[14] Trials were categorised as double blind, unclear or not double blind in three studies,[9,12,14] with the remaining four[11,13,15,21] categorising trials as either double blind or unclear/not double blind.

Table 10 shows characteristics of the 234 meta-analyses and 1973 trials included in the database analysed. The median year of publication was 2000 for meta-analyses and 1989 for trials, whereas the median sample size was 1264 for meta-analyses and 112 for trials. A total of 57 meta-analyses (24.4%) were concerned with conditions related to pregnancy and childbirth, followed by circulatory system conditions (31, 13.3%) and mental health (26, 11.1%). The majority of experimental interventions were pharmacological (162 meta-analyses, 69.2%), whereas placebo or no treatment was the most common comparison intervention (172, 73.5%). A total of 98 meta-analyses (41.9%) analysed a subjectively assessed outcome, followed by all-cause mortality (44, 18.8%), outcomes that are objectively measured, but potentially influenced by patient/clinician judgement (42, 18.0%) and other objectively assessed outcomes (36, 15.4%); 14 meta-analyses (6.0%) contained trials with both objective and subjective outcome measures (e.g. validated and self-reported smoking cessation).

Table 11 summarises the characteristics of trials included in analyses. Information on sequence generation was available for 1207 (61.2%) trials included in 186 meta-analyses, of which 112 meta-analyses containing 944 trials were informative. Sequence generation was assessed as unclear in 769 (63.7%) of these trials, although 306 (25.4%) were assessed as having adequate sequence generation. Percentages were similar for trials included in informative meta-analyses.

Information on allocation concealment was available for most meta-analyses (228, 97.4%) and trials (1796, 91.0%), of which 146 meta-analyses containing 1292 trials were informative. In 1244 (69.3%) trials, allocation concealment was assessed as unclear; 416 (23.2%) trials reported sufficient information to be classed as having adequate allocation concealment. The percentage of trials assessed as having adequate allocation concealment was somewhat higher (29.1%) among those included in informative meta-analyses.

TABLE 8 Characteristics of meta-epidemiological studies contributing to analyses of the influence of reported study design characteristics

Contributing study	Source of systematic reviews/ meta-analyses	Choice of meta-analyses	Study design characteristics examined	No. of meta-analyses (trials)
Als-Nielsen et al.[12,22]	Randomly selected from The Cochrane Library, Issue 2, 2001	Binary outcome and ≥ 5 full-paper trials of which at least one had adequate and one inadequate allocation concealment	Sequence generation, allocation concealment, blinding, intention-to-treat analysis, power calculation	38 (401)
Balk et al.[13]	From four clinical areas (cardiovascular disease, infectious disease, paediatrics, surgery) identified from previous research database, MEDLINE (1966–2000) and The Cochrane Library, Issue 4, 2000	Binary outcome, ≥ 6 trials, significant between-study heterogeneity (OR scale)	27 characteristics including allocation concealment, blinding, intention-to-treat analysis, power calculation, stopping rules, baseline comparability	20 (229)
Contopoulos-Ioannidis et al.[21]	Mental health-related interventions identified from the Mental Health Library, 2002 (Issue 1)	At least one large and one small trial	Trial size, method of randomisation, allocation concealment, blinding	9 (66)
Egger et al.[9]	Meta-analyses from the Cochrane Database of Systematic Reviews that had performed comprehensive literature searches	Outcome measure reported by the largest number of trials	Publication status, language of publication, publication in MEDLINE-indexed journals, allocation sequence generation (subset), allocation concealment, blinding	79 (643)
Kjaergard et al.[15]	In The Cochrane Library, MEDLINE or PubMed with at least one trial with ≥ 1000 patients	Outcome measure described as primary by the review authors or reported by the largest number of trials	Sequence generation, allocation concealment, blinding, description of dropouts and withdrawals	6 (59)
Pildal et al.[14]	Random sample of 38 reviews from The Cochrane Library, Issue 2, 2003, and 32 other reviews from PubMed accessed in 2002	Binary outcome from a meta-analysis presented as the first statistically significant result that supported a conclusion in favour of one of the interventions	Language of publication, sequence generation, allocation concealment, blinding	56 (370)
Schulz et al.[11]	Cochrane Pregnancy and Childbirth Group, ≥ 5 trials containing ≥ 25 events in the control group, at least one trial with and without adequate allocation concealment	The most homogeneous group of interventions	Allocation sequence generation, allocation concealment, blinding, reporting of exclusions	26 (205)

OR, odds ratio.

Information on double blinding was available for all except three trials. However, only 104 meta-analyses (1057 trials) were informative; 77 meta-analyses contained no trials that were double blind, whereas 53 contained only double-blind trials. A total of 929 (47.2%) trials were classified as double blind, compared with 590 (55.8%) trials in informative meta-analyses.

Information on both allocation concealment and blinding was available in 1793 (90.9%) trials contained in 228 (97.4%) meta-analyses, although information on all three study design characteristics was available in 1171 (59.4%) trials contained in 175 (74.8%) meta-analyses.

Table 12 shows associations between the reported study design characteristics, for all trials combined and separately according to the nature of the outcome measure. Trials reporting adequate sequence generation were more likely to report adequate allocation concealment [OR 3.01, 95% confidence interval (CI) 2.20 to 4.12], but there was little association between adequate sequence generation and double blinding (OR 1.03, 95% CI 0.78 to 1.35). However, adequately concealed trials were more likely to be double blind (OR 3.14, 95% CI 2.49 to 3.96).

TABLE 9 Definitions of allocation sequence generation, allocation concealment and blinding used in contributing studies

Study	Sequence generation			Allocation concealment			Blinding		
	Adequate	Unclear	Inadequate	Adequate	Unclear	Inadequate	Double blind	Unclear	Not double blind
Als-Nielsen et al.[1,2,22]	Computer-generated, random number tables, flip of coin, drawing cards or lots, comparable stochastic method	ND	Quasi-randomised (dates, alternation or similar)	Central randomisation (incl. pharmacy controlled), coded identical drug boxes, sealed envelopes, on-site locked computer system or comparable	ND	Open allocation sequence	Described as double blind or at least two key groups (patient/doctor/assessor/ analyst) blinded	ND	Single blind or not blinded
Balk et al.[13]	Did not assess adequacy of sequence generation			Central randomisation; blinded code, coded drug containers; drugs prepared by pharmacy; serially numbered, opaque, sealed envelopes	ND	Random tables, cards, methods using year of birth or registration numbers	Patients and either caregivers or outcome assessors blinded	NA	Any other descriptions not classified as adequate or unclear
Contopoulos-Ioannidis et al.[21]	Computer-generated, random number tables, coin or dice toss, other methods ensuring random order	ND or unclear	Alternation, case records, dates or similar non-random method	Central facility, central pharmacy, with sealed and opaque envelopes	ND or unclear	Any other methods that could not be classified as adequate	Described as double blind, or patients and either outcome assessor or caregiver blinded	NA	Not blinded, single blind, blinding not feasible, ND unclear
[b]Egger et al.[9]	Computer-generated, random number tables or other methods that ensure random order	ND	Alternation, case record numbers, date of birth, etc.	Central randomisation; numbered or coded bottles or containers; drugs prepared by pharmacy; serially numbered, opaque, sealed envelopes; other convincing description implying concealment	ND	Alternation, open random number tables, etc.	Described as double blind	ND	Described as open or similar
Kjaergard et al.[15]	Computer-generated or similar	NA	ND or non-random	Central independent unit, sealed envelopes or similar	NA	ND, open random number table or similar	Described as double blind and used identical placebo or similar	NA	Open (not blind), or described as single blind
Pildal et al.[14]	Computer-generated sequence, random number tables, drawing lots/envelopes, coin toss	ND or unclear	Alternation, case record numbers, date of birth, etc.	Central randomisation; coded drug containers; drugs prepared by central pharmacy, serially numbered, opaque, sealed envelopes; other convincing description implying concealment	ND or approach not falling into other categories	Obvious which treatment the next patient would be allocated (alternation, case record numbers, dates of birth, etc.)	Described as double blind or patients and caregivers reported as blinded, placebo controlled without indication that treatments distinguishable or investigators unblinded[a]	ND or unclear	Not blinded, single blind, did not fit the definition of double blind[a]
Schulz et al.[11]	Computer random number generator, random number tables, coin tossing, shuffling, other random process, minimisation	ND	Non-random	Central randomisation; numbered or coded bottles or containers; drugs prepared by pharmacy; serially numbered, opaque, sealed envelopes; other convincing description implying concealment	ND or approach not falling into other categories	Alternation or allocation by case record number or date of birth	Participants, caregivers and outcome assessors all described as blinded	NA	Descriptions not consistent with definition of double blind, blinding not feasible, ND or unclear

NA, not available; ND, method not described or not mentioned.
a Patient and assessor blinding was not categorised as double blind in the study by Pildal et al.[14]
b Data on sequence generation were available for only a small subset of the study by Egger et al.[9]

TABLE 10 Characteristics of included meta-analyses and trials

Characteristics of meta-analyses and trials	Meta-analyses (*n* = 234)		Trials (*n* = 1973)	
	n	%	*n*	%
Contributing meta-epidemiological study				
Als-Nielsen *et al.*[12,22]	38	16.2	401	20.3
Balk *et al.*[13]	20	8.6	229	11.6
Contopoulos-Ioannidis *et al.*[21]	9	3.9	66	3.4
Egger *et al.*[9]	79	33.8	643	32.6
Kjaergard *et al.*[15]	6	2.6	59	3.0
Pildal *et al.*[14]	56	23.9	370	18.8
Schulz *et al.*[11]	26	11.1	205	10.4
Clinical area according to ICD-10 chapters[30]				
Pregnancy and childbirth (chapter XV, blocks O)	57	24.4	447	22.7
Mental and behavioural (chapter V, F)	26	11.1	302	15.3
Circulatory system (chapter IX, I)	31	13.3	277	14.0
Digestive system (chapter XI, K)	17	7.3	152	7.7
Other factors[a] (chapter XXI, Z)	18	7.7	128	6.5
Respiratory system (chapter X, J)	14	6.0	125	6.3
Other ICD-10 chapters	70	29.9	539	27.3
Unclassified	1	0.4	3	0.2
Type of experimental intervention				
Pharmacological	162	69.2	1418	71.9
Surgical	14	6.0	122	6.2
Psychosocial/behavioural/educational	13	5.6	121	6.1
Other	45	19.2	312	15.8
Type of comparison intervention				
Placebo or no treatment	172	73.5	1438	72.9
Other inactive ('standard care')	16	6.8	152	7.7
Active comparison	44	18.8	368	18.7
Mixture of active and inactive within meta-analysis	2	0.9	15	0.8
Type of outcome measure				
All-cause mortality	44	18.8	364	18.5
Other objective	36	15.4	213	10.8
Objectively measured but influenced by judgement	42	18.0	407	20.6
Subjective	98	41.9	809	41
Mixture of objective and subjective	14	6.0	180	9.1
Year of publication of review[b]/trial				
Median (range)	2000 (1983–2005)		1989 (1948–2002)	
IQR	2000 to 2001		1983 to 1994	
Sample size of meta-analysis/trial				
Median (range)	1264 (72–176,733)		112 (2–82,892)	
IQR	533 to 2582		58 to 267	

ICD-10, *International Classification of Diseases*, 10th Edition.[30]
a   Factors influencing health status and contact with health services (chapter XXI, blocks Z).[30]
b   Year of publication is missing for 26 Schulz reviews that were not retrievable.

TABLE 11 Reported study design characteristics of included trials

Study design characteristic	No. (%) of meta-analyses with information ($n = 234$)	No. (%) of trials with information ($n = 1973$)	No. (%) of informative meta-analyses with information	No. (%) of trials with information included in informative meta-analyses
Adequate sequence generation	186 (79.5)	1207 (61.2)	112 (47.9)	944 (47.8)
Yes		306 (25.4)		248 (26.3)
Unclear		769 (63.7)		598 (63.3)
No		101 (8.4)		67 (7.1)
No/unclear[a]		31 (2.6)		31 (3.3)
Adequate allocation concealment	228 (97.4)	1796 (91.0)	146 (62.4)	1292 (65.5)
Yes		416 (23.2)		376 (29.1)
Unclear		1244 (69.3)		828 (64.1)
No		136 (7.6)		88 (6.8)
Double blind	234 (100.0)	1970 (99.8)	104 (44.4)	1057 (53.6)
Yes		929 (47.2)		590 (55.8)
Unclear		109 (5.5)		63 (6.0)
No		683 (34.7)		249 (23.6)
No/unclear[a]		249 (12.6)		155 (14.7)
Information on both allocation concealment and blinding	228 (97.4)	1793 (90.9)		
Information on all three characteristics	175 (74.8)	1171 (59.4)		

a Some studies did not separate unclear and inadequate sequence generation, or unclear and not double blind.

TABLE 12 Associations between reported study design characteristics

Study characteristic 1	Study characteristic 2	No. (%) of trials			
		All trials	Mortality outcome	Objective outcome	Subjective outcome
Sequence generation	*Allocation concealment*	1171	157	368	646
Adequate	Adequate	91 (7.8)	16 (10.2)	32 (8.7)	43 (6.7)
Adequate	Inadequate/unclear	182 (15.5)	25 (15.9)	64 (17.4)	93 (14.4)
Inadequate/unclear	Adequate	128 (10.9)	15 (9.6)	45 (12.2)	68 (10.5)
Inadequate/unclear	Inadequate/unclear	770 (65.8)	101 (64.3)	227 (61.7)	442 (68.4)
OR (95% CI)		3.01 (2.20 to 4.12)	4.31 (1.88 to 9.88)	2.52 (1.48 to 4.29)	3.01 (1.93 to 4.68)
Sequence generation	*Blinding*	1171	157	368	646
Adequate	Double blind	127 (10.8)	21 (13.4)	44 (12.0)	62 (9.6)
Adequate	Not double blind/unclear	146 (12.5)	20 (12.7)	52 (14.1)	74 (11.5)
Inadequate/unclear	Double blind	412 (35.2)	53 (33.8)	127 (34.5)	232 (35.9)
Inadequate/unclear	Not double blind/unclear	486 (41.5)	63 (40.1)	145 (39.4)	278 (43.0)
OR (95% CI)		1.03 (0.78 to 1.35)	1.25 (0.61 to 2.55)	0.97 (0.61 to 1.54)	1.00 (0.69 to 1.47)
Allocation concealment	*Blinding*	1793	328	550	915
Adequate	Double blind	283 (15.8)	65 (19.8)	93 (16.9)	125 (13.7)
Adequate	Not double blind/unclear	133 (7.4)	30 (9.1)	45 (8.2)	58 (6.3)
Inadequate/unclear	Double blind	556 (31.0)	108 (32.9)	159 (28.9)	289 (31.6)
Inadequate/unclear	Not double blind/unclear	821 (45.8)	125 (38.1)	253 (46.0)	443 (48.4)
OR (95% CI)		3.14 (2.49 to 3.96)	2.51 (1.52 to 4.15)	3.29 (2.19 to 4.94)	3.30 (2.34 to 4.66)

Associations between reported study design characteristics were similar for the different types of outcome measure. *Table 13* shows the number of trials assessed with each of the eight possible combinations of the study design characteristics, overall and according to the type of outcome measure. Only 60 (5.1%) trials were assessed as at low risk of bias for all three characteristics, whereas 453 (38.7%) were assessed as at high risk of bias for all three characteristics.

Influence of reported study design characteristics on intervention effect estimates: univariable analyses of individual characteristics

Figure 3 and *Table 14* present results from univariable analyses of the influence of reported study design characteristics on intervention effect estimates, both overall and separately according to type of outcome measure. Compared with *Figure 3*, *Table 14* additionally includes 95% CrIs for the variance parameters κ and φ and displays the numbers of trials, and trials at high risk of bias, included in analyses. Overall, intervention effect estimates were exaggerated by an average of 11% in trials with inadequate or unclear sequence generation (ROR 0.89, 95% CrI 0.82 to 0.96), and between-trial heterogeneity was higher among such trials ($\kappa = 0.16$, 95% CrI 0.03 to 0.27). When analyses were stratified according to the type of outcome measure, the average effect of inadequate or unclear sequence generation appeared greatest for subjective outcomes [ROR 0.83, 95% CrI 0.74 to 0.94, posterior probability (PPr) that $ROR_{subjective} < ROR_{mortality} = 0.73$], and the increase in between-trial heterogeneity was also greatest for such outcomes ($\kappa = 0.20$, CrI 0.03 to 0.32, PPr that $\kappa_{subjective} > \kappa_{mortality} = 0.78$). In contrast, there was little evidence that inadequate or unclear sequence generation was associated with exaggeration of intervention effects for all-cause mortality (ROR 0.89, 95% CrI 0.75 to 1.05) or for other objective outcomes (ROR 0.99, 95% CrI 0.84 to 1.16). For all types of outcome measure there was only limited between-meta-analysis heterogeneity in mean bias (estimated φ between 0.04 and 0.07).

Overall, intervention effect estimates were exaggerated by 7% in trials with inadequate or unclear allocation concealment (ROR 0.93, 95% CrI 0.87 to 0.99), and there was evidence that between-trial heterogeneity was increased for such studies ($\kappa = 0.12$, 95% CrI 0.02 to 0.23). The influence of inadequate or unclear allocation concealment appeared greatest among meta-analyses with a subjectively assessed outcome measure (ROR 0.85, 95% CrI 0.75 to 0.95, PPr that $ROR_{subjective} < ROR_{mortality} = 0.97$; $\kappa = 0.20$, 95% CrI 0.02 to 0.33, PPr that $\kappa_{subjective} > \kappa_{mortality} = 0.85$). In contrast, the average effect of inadequate or unclear allocation concealment was close to the null for meta-analyses with mortality (ROR 0.98, 95% CrI 0.88 to 1.10) and other objective outcomes

TABLE 13 Number of trials with each combination of the three reported study design characteristics, overall and according to type of outcome measure

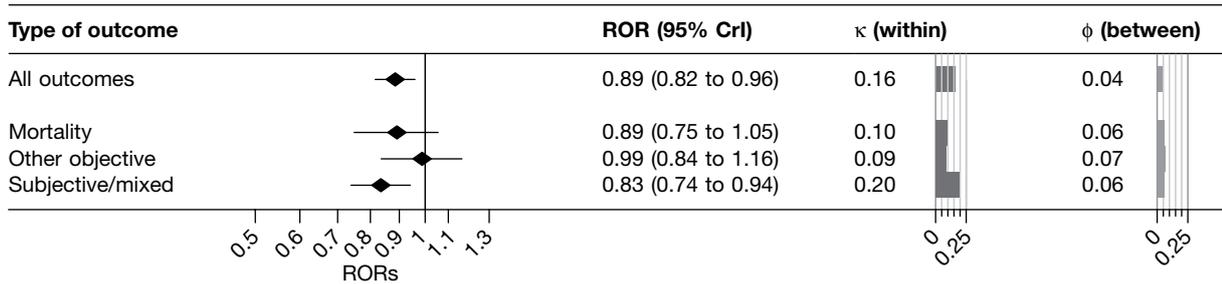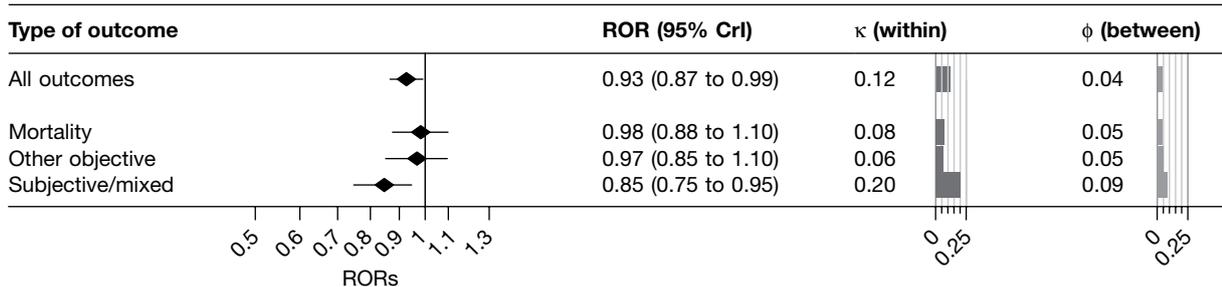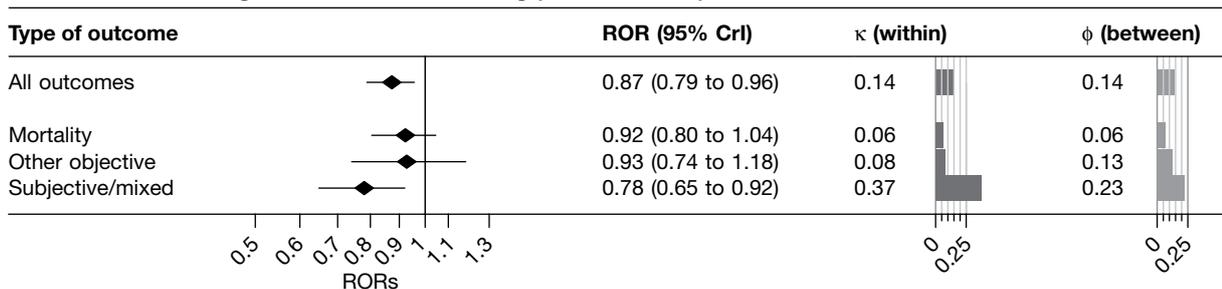| Study design characteristics			Number of trials			
Sequence generation	Allocation concealment	Blinding	All	Mortality	Objective	Subjective
Adequate	Adequate	Double blind	60	12	22	26
Adequate	Adequate	Not double blind[a]	31	4	10	17
Adequate	Inadequate[a]	Double blind	67	9	22	36
Adequate	Inadequate[a]	Not double blind[a]	115	16	42	57
Inadequate[a]	Adequate	Double blind	95	9	35	51
Inadequate[a]	Adequate	Not double blind[a]	33	6	10	17
Inadequate[a]	Inadequate[a]	Double blind	317	44	92	181
Inadequate[a]	Inadequate[a]	Not double blind[a]	453	57	135	261
Total			1171	157	368	646

a   Or unclear.

Inadequate or unclear generation of randomisation sequence (vs adequate)

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	0.89 (0.82 to 0.96)	0.16	0.04
Mortality	0.89 (0.75 to 1.05)	0.10	0.06
Other objective	0.99 (0.84 to 1.16)	0.09	0.07
Subjective/mixed	0.83 (0.74 to 0.94)	0.20	0.06

Inadequate or unclear allocation concealment (vs adequate)

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	0.93 (0.87 to 0.99)	0.12	0.04
Mortality	0.98 (0.88 to 1.10)	0.08	0.05
Other objective	0.97 (0.85 to 1.10)	0.06	0.05
Subjective/mixed	0.85 (0.75 to 0.95)	0.20	0.09

Lack of double blinding or unclear double blinding (vs double blind)

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	0.87 (0.79 to 0.96)	0.14	0.14
Mortality	0.92 (0.80 to 1.04)	0.06	0.06
Other objective	0.93 (0.74 to 1.18)	0.08	0.13
Subjective/mixed	0.78 (0.65 to 0.92)	0.37	0.23

FIGURE 3 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, according to the type of outcome measure: univariable analyses based on all available data. Random sequence generation: 944 trials from 112 informative meta-analyses; allocation concealment: 1292 trials from 146 informative meta-analyses; blinding: 1057 trials from 104 informative meta-analyses.

(ROR 0.97, 95% CrI, 0.85 to 1.10). Estimates of both between-trial and between-meta-analyses heterogeneity in bias were lower for such outcomes than for subjectively assessed outcomes.

Lack of, or unclear, double blinding was associated with an average 13% exaggeration of intervention effects (ROR 0.87, 95% CrI 0.79 to 0.96). There was evidence that between-trial heterogeneity was increased for such studies ($\kappa = 0.14$, 95% CrI 0.02 to 0.30), and that average bias varied between meta-analyses ($\varphi = 0.14$, 95% CrI 0.03 to 0.28). Average bias (ROR 0.78, 95% CrI 0.65 to 0.92), increased between-trial heterogeneity ($\kappa = 0.37$, 95% CrI 0.19 to 0.53) and between-meta-analysis heterogeneity in average bias ($\varphi = 0.23$, 95% CrI 0.04 to 0.44) all appeared greatest for meta-analyses assessing subjective outcomes (PPr $\text{ROR}_{subjective} < \text{ROR}_{mortality} = 0.94$, PPr $\kappa_{subjective} > \kappa_{mortality} = 0.99$, PPr $\varphi_{subjective} > \varphi_{mortality} = 0.90$). Among meta-analyses with subjectively assessed outcomes, the influence of lack of blinding appeared greater than the influence of inadequate or unclear sequence generation or allocation concealment.

Results from univariable analyses of the influence of inadequate or unclear allocation concealment and lack of double blinding restricted to the 228 meta-analyses (1793 trials) that contained information on both of these characteristics were similar (*Table 15*) to those presented in *Figure 3* and *Table 14*. We also repeated the univariable analyses in a data set restricted to

TABLE 14 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, according to type of outcome measure: univariable analyses based on all available data

Study design characteristic and outcome	No. of meta-analyses	No. of trials	Contributing meta-analyses[a]	Contributing trials	No. (%) of trials at high risk of bias[b]	ROR	95% CrI	κ	95% CrI	ϕ	95% CrI	V_0
Inadequate or unclear sequence generation (vs adequate)												
All	186	1207	112 (57)	944	696 (73.7)	0.89	0.82 to 0.96	0.16	0.03 to 0.27	0.04	0.01 to 0.14	0.002
Mortality	30	163	16 (12)	129	92 (71.3)	0.89	0.75 to 1.05	0.10	0.01 to 0.29	0.06	0.01 to 0.25	0.008
Objective	66	387	47 (19)	328	234 (71.3)	0.99	0.84 to 1.16	0.09	0.01 to 0.34	0.07	0.01 to 0.25	0.007
Subjective	90	657	49 (26)	487	370 (76.0)	0.83	0.74 to 0.94	0.20	0.03 to 0.32	0.06	0.01 to 0.22	0.004
Inadequate or unclear allocation concealment (vs adequate)												
All	228	1796	146 (80)	1292	916 (70.9)	0.93	0.87 to 0.99	0.12	0.02 to 0.23	0.04	0.01 to 0.13	0.001
Mortality	44	328	32 (15)	268	183 (68.3)	0.98	0.88 to 1.10	0.08	0.01 to 0.21	0.05	0.01 to 0.18	0.003
Objective	76	551	45 (21)	372	253 (68.0)	0.97	0.85 to 1.10	0.06	0.01 to 0.26	0.05	0.01 to 0.24	0.004
Subjective	108	917	69 (44)	652	480 (73.6)	0.85	0.75 to 0.95	0.20	0.02 to 0.33	0.09	0.01 to 0.29	0.003
Lack of double blinding or unclear double blinding (vs double blind)												
All	234	1970	104 (66)	1057	467 (44.2)	0.87	0.79 to 0.96	0.14	0.02 to 0.30	0.14	0.03 to 0.28	0.002
Mortality	44	364	25 (14)	245	109 (44.5)	0.92	0.80 to 1.04	0.06	0.01 to 0.20	0.06	0.01 to 0.22	0.004
Objective	78	619	28 (17)	282	120 (42.6)	0.93	0.74 to 1.18	0.08	0.01 to 0.38	0.13	0.01 to 0.50	0.014
Subjective	112	987	51 (35)	530	238 (44.9)	0.78	0.65 to 0.92	0.37	0.19 to 0.53	0.23	0.04 to 0.44	0.008

a   Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
b   High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.

175 meta-analyses (1171 trials) that had information on all three study design characteristics (sequence generation, allocation concealment and blinding). The results were similar (*Table 16*), although the influence of inadequate allocation concealment appeared somewhat greater in analyses restricted to the 88 informative meta-analyses (811 trials) that had information on all three characteristics.

Influence of reported study design characteristics: univariable analyses of combinations of characteristics

We conducted further univariable analyses using variables defined based on combinations of the study design characteristics. *Figure 4* and *Table 16* show the estimated effects of any risk of selection bias (inadequate or unclear sequence generation or allocation concealment, compared with other trials) on intervention effects and heterogeneity, based on 53 informative meta-analyses containing 534 trials, of which 89 (17%) were assessed as being at low risk of selection bias. Risk of selection bias was associated with an average 11% exaggeration of intervention effect estimates (ROR 0.89, 95% CrI 0.78 to 1.00) and with increased between-trial heterogeneity ($\kappa = 0.12$, 95% CrI 0.02 to 0.27). Average effects did not differ substantially according to type of outcome measure. Between-meta-analysis heterogeneity was highest for objective outcomes ($\varphi = 0.20$, 95% CrI 0.02 to 0.68).

Only 37 informative meta-analyses [409 trials, of which 58 (14%) were assessed as being at low risk of bias] contributed to the analysis of any risk of bias (inadequate or unclear sequence generation or allocation concealment, or lack of or unclear double blinding, compared with all other trials); *Figure 4* and *Table 16* show estimated effects of any risk of bias from one of the three characteristics (compared with low risk of bias from all three). Any risk of bias was associated with an average 21% exaggeration of intervention effect estimates (ROR 0.79, 95% CrI 0.64 to 0.92). The numbers of informative meta-analyses and trials included in analyses stratified by type of outcome measure were small, but the influence of any risk of bias appeared smallest for all-cause mortality outcomes.

A total of 104 informative meta-analyses [990 trials, of which 259 (26%) were assessed as at low risk of bias] contributed to analyses of the influence of inadequate or unclear allocation concealment or lack of double blinding (compared with adequate allocation concealment and presence of double blinding). *Figure 4* and *Table 15* show that intervention effects from trials at high risk of bias according to this variable were exaggerated by an average of 12% (ROR 0.88, 95% CrI 0.81 to 0.95). The ROR was closer to 1 (0.95) for all-cause mortality outcomes than for other objective or subjective outcomes (0.84, 95% CrI 0.69 to 1.00 and 0.83, 95% CrI 0.73 to 0.93 respectively). The increase in between-trial heterogeneity appeared greatest for subjective outcomes ($\kappa = 0.17$, 95% CrI 0.02 to 0.31).

Univariable analyses: comparison with meta-meta-analytic approach

Table 17 displays results of univariable analyses, based on all available data, conducted using the meta-meta-analytic approach[8] employed in previous work.[16] Using this approach, we estimated the ROR and the between-meta-analysis standard deviation φ associated with each study design characteristic. Increases in between-trial variability are not estimated using this approach.

For inadequate or unclear sequence generation, inadequate or unclear allocation concealment and lack of or unclear double blinding, estimated RORs are broadly consistent with those from the hierarchical bias models displayed in *Figure 3*. For each study design characteristic, the exaggeration of intervention effect estimates is greater for meta-analyses with subjectively assessed outcomes than for meta-analyses with all-cause mortality or other objectively assessed outcomes. Consistent with the small estimated values of φ in the hierarchical models, the between-meta-analysis standard deviation was estimated as zero for each of these characteristics

TABLE 15 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, according to type of outcome measure: univariable analyses based on trials for which assessments of both allocation concealment and double blinding were available

Study design characteristic and outcome	No. of meta-analyses	No. of trials	Contributing meta-analyses[a]	Contributing trials	No. (%) of trials at high risk of bias[b]	ROR	95% CrI	κ	95% CrI	φ	95% CrI
Inadequate or unclear allocation concealment (vs adequate)											
All	228	1793	146 (80)	1291	915 (70.9)	0.93	0.87 to 0.99	0.11	0.02 to 0.22	0.04	0.01 to 0.13
Mortality	44	328	32 (15)	268	183 (68.3)	0.98	0.88 to 1.10	0.08	0.01 to 0.21	0.05	0.01 to 0.18
Objective	76	550	45 (21)	372	253 (68.0)	0.97	0.85 to 1.10	0.06	0.01 to 0.26	0.05	0.01 to 0.21
Subjective	108	915	69 (44)	651	479 (73.6)	0.84	0.75 to 0.94	0.20	0.03 to 0.33	0.09	0.01 to 0.30
Lack of double blinding or unclear double blinding (vs double blind)											
All	228	1793	101 (62)	977	432 (44.2)	0.87	0.78 to 0.96	0.13	0.02 to 0.29	0.13	0.02 to 0.26
Mortality	44	328	25 (13)	215	99 (46.0)	0.92	0.80 to 1.05	0.06	0.01 to 0.20	0.07	0.01 to 0.24
Objective	76	550	27 (17)	259	108 (41.7)	0.88	0.71 to 1.11	0.07	0.01 to 0.38	0.13	0.01 to 0.49
Subjective	108	915	49 (32)	503	225 (44.7)	0.79	0.65 to 0.93	0.33	0.08 to 0.49	0.21	0.04 to 0.41
Inadequate or unclear allocation concealment or not double blind (vs adequate allocation concealment and double blind)											
All	228	1793	104 (55)	990	731 (73.8)	0.88	0.81 to 0.95	0.12	0.02 to 0.22	0.05	0.01 to 0.14
Mortality	44	328	25 (13)	220	157 (71.4)	0.95	0.84 to 1.06	0.08	0.01 to 0.22	0.05	0.01 to 0.19
Objective	76	550	30 (16)	268	191 (71.3)	0.84	0.69 to 1.00	0.07	0.01 to 0.33	0.07	0.01 to 0.33
Subjective	108	915	49 (26)	502	383 (76.3)	0.83	0.73 to 0.93	0.17	0.02 to 0.31	0.06	0.01 to 0.23

a   Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
b   High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.

TABLE 16 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, according to type of outcome measure: univariable analyses based on trials for which assessments of all three characteristics were available

Study design characteristic and outcome	No. of meta-analyses	No. of trials	Contributing meta-analyses[a]	Contributing trials	No. (%) of trials at high risk of bias[b]	ROR	95% CrI	κ	95% CrI	φ	95% CrI
Inadequate or unclear sequence generation (vs adequate)											
All	175	1171	104 (54)	911	676 (74.2)	0.88	0.81 to 0.96	0.17	0.03 to 0.27	0.04	0.01 to 0.15
Mortality	27	157	15 (11)	122	88 (72.1)	0.88	0.73 to 1.05	0.11	0.01 to 0.32	0.06	0.01 to 0.26
Objective	62	368	42 (18)	310	223 (71.9)	0.99	0.84 to 1.18	0.08	0.01 to 0.32	0.07	0.01 to 0.27
Subjective	86	646	47 (25)	479	365 (76.2)	0.83	0.73 to 0.93	0.20	0.03 to 0.32	0.06	0.01 to 0.22
Inadequate or unclear allocation concealment (vs adequate)											
All	175	1171	88 (45)	811	605 (74.6)	0.88	0.80 to 0.96	0.13	0.02 to 0.26	0.05	0.01 to 0.18
Mortality	27	157	15 (5)	118	89 (75.4)	0.97	0.80 to 1.18	0.09	0.01 to 0.31	0.07	0.01 to 0.32
Objective	62	368	31 (15)	257	180 (70.0)	0.93	0.80 to 1.09	0.06	0.01 to 0.24	0.06	0.01 to 0.26
Subjective	86	646	42 (25)	436	336 (77.1)	0.79	0.67 to 0.90	0.22	0.03 to 0.36	0.10	0.01 to 0.36
Lack of double blinding or unclear double blinding (vs double blind)											
All	175	1171	60 (36)	592	251 (42.4)	0.83	0.72 to 0.95	0.26	0.05 to 0.42	0.16	0.03 to 0.33
Mortality	27	157	9 (3)	74	29 (39.2)	1.06	0.82 to 1.44	0.11	0.01 to 0.47	0.08	0.01 to 0.41
Objective	62	368	17 (9)	165	73 (44.2)	0.89	0.64 to 1.26	0.10	0.01 to 0.49	0.22	0.02 to 0.79
Subjective	86	646	34 (24)	353	149 (42.2)	0.73	0.57 to 0.89	0.33	0.09 to 0.50	0.22	0.05 to 0.42
Inadequate or unclear sequence generation or allocation concealment (vs adequate sequence generation and allocation concealment)											
All	175	1171	53 (22)	534	445 (83.3)	0.89	0.78 to 1.00	0.12	0.02 to 0.27	0.06	0.01 to 0.22
Mortality	27	157	10 (4)	79	65 (82.3)	0.94	0.74 to 1.15	0.08	0.01 to 0.29	0.08	0.01 to 0.38
Objective	62	368	19 (10)	176	144 (81.8)	0.82	0.57 to 1.10	0.15	0.01 to 0.61	0.20	0.02 to 0.68
Subjective	86	646	24 (8)	279	236 (84.6)	0.85	0.70 to 1.01	0.15	0.01 to 0.32	0.08	0.01 to 0.34
Inadequate or unclear sequence generation or allocation concealment or not double blind (vs adequate sequence generation and allocation concealment and double blind)											
All	175	1171	37 (14)	409	351 (85.8)	0.79	0.64 to 0.92	0.12	0.01 to 0.28	0.12	0.02 to 0.40
Mortality	27	157	7 (3)	65	55 (84.6)	0.94	0.72 to 1.19	0.09	0.01 to 0.32	0.08	0.01 to 0.38
Objective	62	368	14 (6)	139	117 (84.2)	0.63	0.42 to 0.98	0.24	0.02 to 0.77	0.23	0.02 to 0.82
Subjective	86	646	16 (5)	205	179 (87.3)	0.71	0.52 to 0.89	0.12	0.01 to 0.31	0.16	0.02 to 0.55

a Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
b High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.
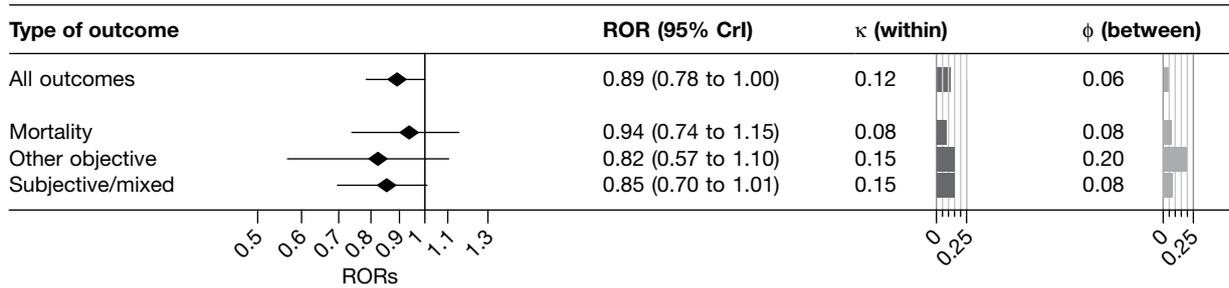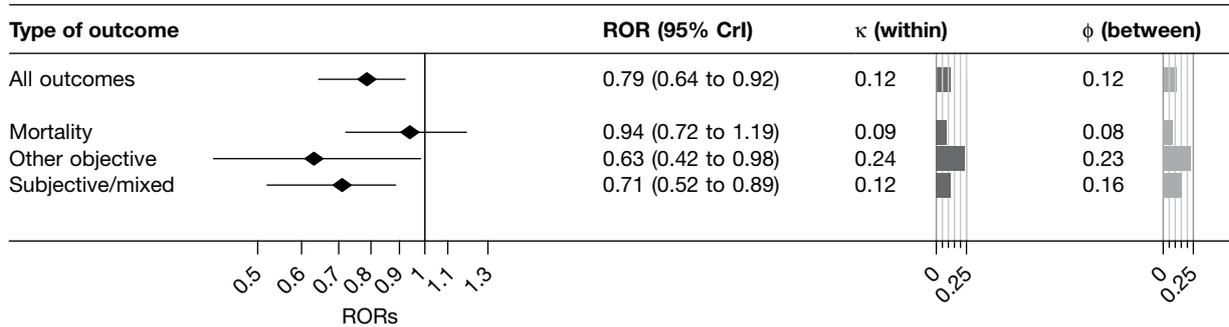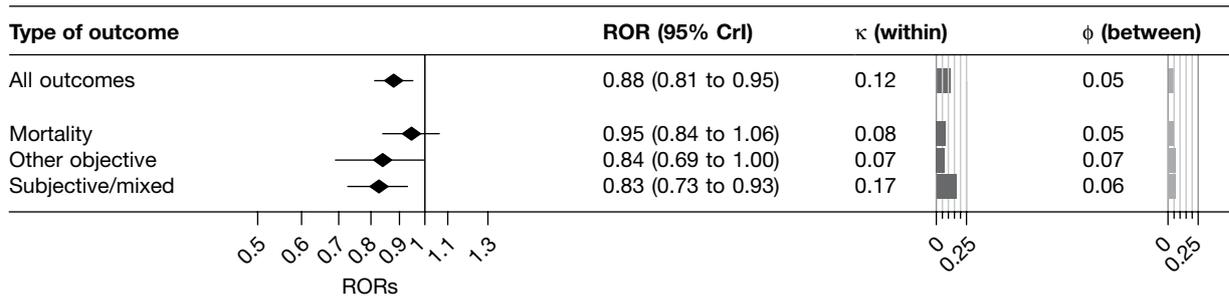
Inadequate or unclear sequence generation or allocation concealment (vs both adequate)

Type of outcome	RORs	ROR (95% CrI)	κ (within)		ϕ (between)	
All outcomes		0.89 (0.78 to 1.00)	0.12		0.06	
Mortality		0.94 (0.74 to 1.15)	0.08		0.08	
Other objective		0.82 (0.57 to 1.10)	0.15		0.20	
Subjective/mixed		0.85 (0.70 to 1.01)	0.15		0.08	

RORs axis: 0.5 0.6 0.7 0.8 0.9 1 1.1 1.3 — κ axis: 0 0.25 — ϕ axis: 0 0.25

Inadequate or unclear for any of the three domains (vs all three adequate)

Type of outcome	RORs	ROR (95% CrI)	κ (within)		ϕ (between)	
All outcomes		0.79 (0.64 to 0.92)	0.12		0.12	
Mortality		0.94 (0.72 to 1.19)	0.09		0.08	
Other objective		0.63 (0.42 to 0.98)	0.24		0.23	
Subjective/mixed		0.71 (0.52 to 0.89)	0.12		0.16	

RORs axis: 0.5 0.6 0.7 0.8 0.9 1 1.1 1.3 — κ axis: 0 0.25 — ϕ axis: 0 0.25

Inadequate or unclear allocation concealment or not double blind (vs both adequate)

Type of outcome	RORs	ROR (95% CrI)	κ (within)		ϕ (between)	
All outcomes		0.88 (0.81 to 0.95)	0.12		0.05	
Mortality		0.95 (0.84 to 1.06)	0.08		0.05	
Other objective		0.84 (0.69 to 1.00)	0.07		0.07	
Subjective/mixed		0.83 (0.73 to 0.93)	0.17		0.06	

RORs axis: 0.5 0.6 0.7 0.8 0.9 1 1.1 1.3 — κ axis: 0 0.25 — ϕ axis: 0 0.25

FIGURE 4 Estimated RORs and effects on heterogeneity associated with combinations of study design characteristics: univariable analyses according to type of outcome measure. Analyses based on subsets of trials with assessments of all three study design characteristics (upper and middle panels), and of allocation concealment and blinding (lower panel). Upper panel: inadequate or unclear sequence generation or allocation concealment (any risk of selection bias); middle panel: inadequate or unclear sequence generation or allocation concealment or lack of or unclear double blinding (any risk of bias); lower panel: inadequate or unclear allocation concealment or lack of or unclear double blinding.

in meta-analyses with all-cause mortality outcomes and also, for sequence generation and allocation concealment, in trials with other objectively assessed outcomes. Between-trial variability in bias was greatest for lack of or unclear double blinding, in meta-analyses with other objective or subjective outcomes. Results from analyses of combined study design characteristics were also broadly consistent with those from the hierarchical bias models.

Influence of reported study design characteristics: multivariable analyses

Table 18 presents results from multivariable analyses of the influence of inadequate or unclear allocation concealment and lack of or unclear double blinding, based on the 169 informative meta-analyses (1456 trials) in which both characteristics were assessed. Results from models without interaction terms are displayed in *Figure 5*: estimated RORs were similar to, or modestly attenuated compared with, the univariable analyses presented in *Table 15*. Estimated effects on heterogeneity were also modestly attenuated. Results from the model including interaction terms are displayed in *Figure 6*. RORs for the interaction between inadequate or unclear allocation

TABLE 17 Estimated RORs and between-meta-analysis heterogeneity in mean bias associated with reported study design characteristics: univariable analyses using meta-meta-analytic approach[16]

Study design characteristic and outcome	No. of trials	ROR (95% CI)	φ
Inadequate or unclear sequence generation (vs adequate)			
All	944	0.94 (0.89 to 1.00)	0.00
Mortality	129	0.91 (0.78 to 1.05)	0.00
Objective	328	1.01 (0.89 to 1.15)	0.00
Subjective	487	0.87 (0.79 to 0.96)	0.08
Inadequate or unclear allocation concealment (vs adequate)			
All	1292	0.93 (0.88 to 0.99)	0.08
Mortality	268	0.99 (0.90 to 1.09)	0.00
Objective	372	0.98 (0.88 to 1.09)	0.00
Subjective	652	0.85 (0.76 to 0.95)	0.19
Lack of double blinding or unclear double blinding (vs double blind)			
All	1057	0.86 (0.79 to 0.94)	0.21
Mortality	245	0.90 (0.80 to 1.00)	0.00
Objective	282	0.97 (0.77 to 1.21)	0.32
Subjective	530	0.78 (0.67 to 0.90)	0.27
Inadequate or unclear sequence generation or allocation concealment (vs adequate sequence generation and allocation concealment)			
All	534	0.95 (0.88 to 1.03)	0.04
Mortality	79	0.97 (0.83 to 1.13)	0.00
Objective	176	0.91 (0.72 to 1.14)	0.03
Subjective	279	0.91 (0.80 to 1.03)	0.10
Inadequate or unclear sequence generation or allocation concealment or not double blind (vs adequate sequence generation and allocation concealment and double blind)			
All	409	0.81 (0.71 to 0.93)	0.14
Mortality	65	0.96 (0.82 to 1.13)	0.00
Objective	139	0.68 (0.47 to 0.96)	0.00
Subjective	205	0.72 (0.57 to 0.91)	0.25
Inadequate or unclear allocation concealment or not double blind (vs adequate allocation concealment and double blind)			
All	990	0.91 (0.87 to 0.96)	0.00
Mortality	220	0.92 (0.83 to 1.02)	0.00
Objective	268	0.86 (0.74 to 1.00)	0.00
Subjective	502	0.83 (0.75 to 0.93)	0.13

concealment and lack of double blinding were close to 1, with wide CIs. Therefore, these results are consistent with the effects of these two characteristics being multiplicative.

Results from multivariable analyses of the influence of all three study design characteristics are presented in *Table 19* and displayed in *Figure 7*. Estimated RORs for each study design characteristic were of similar magnitudes to those in the univariable analyses presented in *Table 16*. For inadequate or unclear sequence generation or allocation concealment, estimated increases in between-trial heterogeneity (quantified by κ) were smaller in multivariable analyses than in the corresponding univariable analyses (see *Table 16*). Estimates of between-meta-analysis variability in average bias changed little compared with univariable analyses.

TABLE 18 Estimated RORs and effects on heterogeneity associated with inadequate or unclear allocation concealment and lack of or unclear double blinding: results from multivariable analyses[a]

Model, study design characteristic and outcome	ROR	95% CrI	κ	95% CrI	φ	95% CrI
Model without interaction						
Inadequate or unclear allocation concealment (vs adequate)						
All	0.93	0.87 to 1.00	0.08	0.01 to 0.20	0.05	0.01 to 0.14
Mortality	1.00	0.89 to 1.13	0.06	0.01 to 0.20	0.05	0.01 to 0.18
Objective	0.97	0.84 to 1.13	0.06	0.01 to 0.25	0.05	0.01 to 0.21
Subjective	0.85	0.76 to 0.96	0.07	0.01 to 0.27	0.06	0.01 to 0.22
Lack of double blinding or unclear double blinding (vs double blind)						
All	0.88	0.79 to 0.97	0.12	0.02 to 0.28	0.13	0.03 to 0.28
Mortality	0.92	0.80 to 1.06	0.06	0.01 to 0.20	0.07	0.01 to 0.24
Objective	0.90	0.71 to 1.15	0.08	0.01 to 0.38	0.16	0.01 to 0.54
Subjective	0.82	0.68 to 0.96	0.30	0.04 to 0.48	0.19	0.03 to 0.40
Implied average bias in trials with high risk of bias for both characteristics						
All	0.83	0.74 to 0.92				
Mortality	0.92	0.78 to 1.09				
Objective	0.87	0.68 to 1.12				
Subjective	0.70	0.57 to 0.84				
Model including interaction terms						
Inadequate or unclear allocation concealment (vs adequate), in double-blind trials						
All	0.90	0.83 to 0.98	0.07	0.01 to 0.19	0.05	0.01 to 0.14
Mortality	0.98	0.84 to 1.13	0.06	0.01 to 0.20	0.05	0.01 to 0.19
Objective	0.86	0.70 to 1.02	0.06	0.01 to 0.24	0.05	0.01 to 0.21
Subjective	0.85	0.73 to 0.96	0.08	0.01 to 0.25	0.07	0.01 to 0.22
Lack of double blinding or unclear double blinding (vs double blind), in adequately concealed trials						
All	0.84	0.73 to 0.95	0.10	0.01 to 0.26	0.11	0.01 to 0.26
Mortality	0.87	0.69 to 1.06	0.06	0.01 to 0.20	0.06	0.01 to 0.24
Objective	0.72	0.52 to 1.01	0.08	0.01 to 0.38	0.15	0.01 to 0.56
Subjective	0.85	0.67 to 1.05	0.16	0.01 to 0.41	0.09	0.01 to 0.29
Interaction						
All	1.08	0.95 to 1.24	0.10	0.01 to 0.27	0.07	0.01 to 0.24
Mortality	1.10	0.85 to 1.43	0.06	0.01 to 0.21	0.05	0.01 to 0.22
Objective	1.30	1.00 to 1.78	0.07	0.01 to 0.28	0.06	0.01 to 0.26
Subjective	0.95	0.74 to 1.21	0.20	0.01 to 0.43	0.23	0.02 to 0.46
Implied average bias in trials with high risk of bias for both characteristics						
All	0.82	0.73 to 0.91				
Mortality	0.93	0.78 to 1.10				
Objective	0.81	0.63 to 1.05				
Subjective	0.68	0.55 to 0.83				

a   Numbers of contributing trials/meta-analyses were 1456/169 (all outcomes), 294/36 (all-cause mortality), 424/53 (other objective) and 738/80 (subjective).
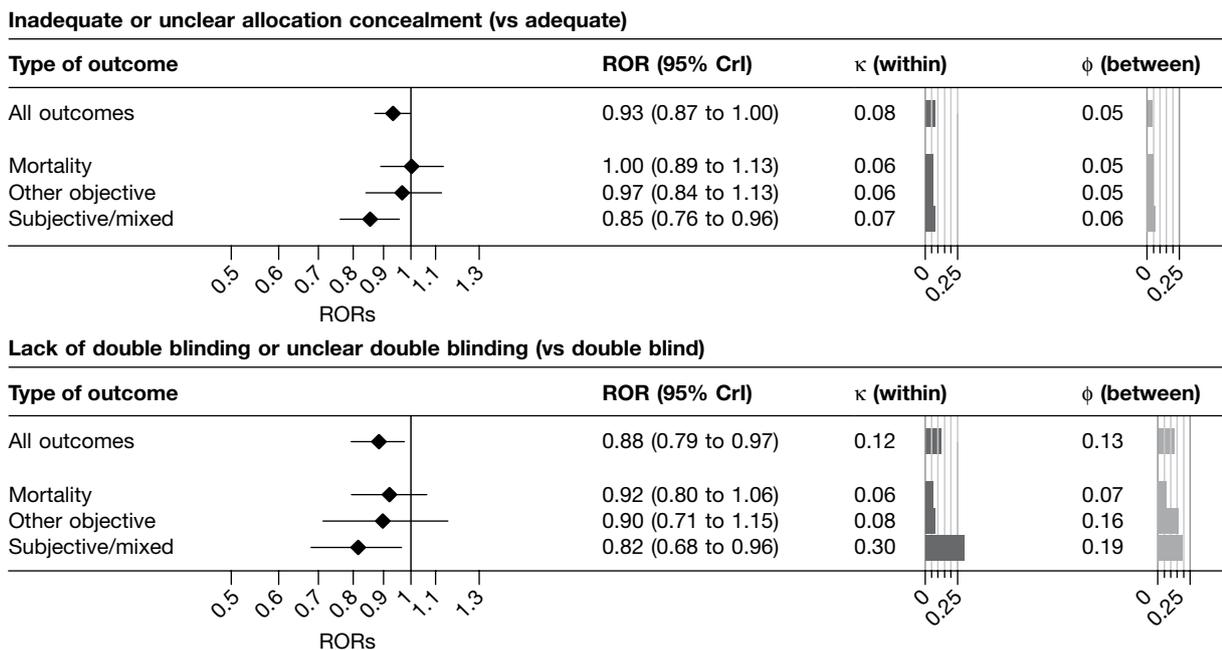
Inadequate or unclear allocation concealment (vs adequate)

Type of outcome		ROR (95% CrI)	κ (within)		φ (between)	
All outcomes		0.93 (0.87 to 1.00)	0.08		0.05	
Mortality		1.00 (0.89 to 1.13)	0.06		0.05	
Other objective		0.97 (0.84 to 1.13)	0.06		0.05	
Subjective/mixed		0.85 (0.76 to 0.96)	0.07		0.06	

RORs

Lack of double blinding or unclear double blinding (vs double blind)

Type of outcome		ROR (95% CrI)	κ (within)		φ (between)	
All outcomes		0.88 (0.79 to 0.97)	0.12		0.13	
Mortality		0.92 (0.80 to 1.06)	0.06		0.07	
Other objective		0.90 (0.71 to 1.15)	0.08		0.16	
Subjective/mixed		0.82 (0.68 to 0.96)	0.30		0.19	

RORs

FIGURE 5 Estimated RORs and effects on heterogeneity from multivariable analyses, according to type of outcome measure. Analyses based on trials with assessments of both allocation concealment and blinding. Upper panel, influence of inadequate or unclear versus adequate allocation concealment, adjusted for blinding; lower panel, influence of lack of double blinding or unclear blinding status versus double blinding, adjusted for allocation concealment.

Analyses according to type of intervention and clinical area.

Results from univariable analyses of the influence of the three study design characteristics according to clinical area are shown in *Table 20*. For pregnancy and childbirth – the clinical area contributing most meta-analyses to the combined data set – RORs were further from 1 than in the analyses of the whole data set displayed in *Figure 3* and *Table 14*, whereas estimates of effects on heterogeneity were broadly consistent with analyses of the whole data set. For the other two clinical areas, RORs were attenuated towards 1. Only small numbers of meta-analyses contributed to estimation of κ, but estimated values of κ and φ were generally smaller for circulatory system meta-analyses than for the other two clinical areas.

Table 21 displays results of univariable analyses restricted to pharmacological and surgical interventions. The majority of meta-analyses included in the full data set addressed pharmacological interventions; it was therefore unsurprising that overall results restricted to such interventions were consistent with those from the full data set. For surgical interventions, the influence of inadequate or unclear sequence generation and allocation concealment was estimated from only six and nine meta-analyses respectively: CIs were too wide to allow substantive conclusions to be drawn.

Analyses of meta-analyses comparing two active interventions

For meta-analyses comparing two active interventions we estimated increases in between-trial (within-meta-analysis) heterogeneity in trials with, compared with trials without, a study design characteristic of interest (among meta-analyses containing at least two trials with and without the characteristic of interest). Based on eight meta-analyses containing 84 trials, the between-trial standard deviation (corresponding to κ in previous analyses) for trials with inadequate or unclear (compared with adequate) sequence generation was 0.35 (95% CrI 0.02 to 0.88), similar to the estimate for inadequate or unclear (compared with adequate) allocation concealment (0.36, 95% CrI 0.02 to 0.88, based on six meta-analyses containing 52 trials). The estimated increase
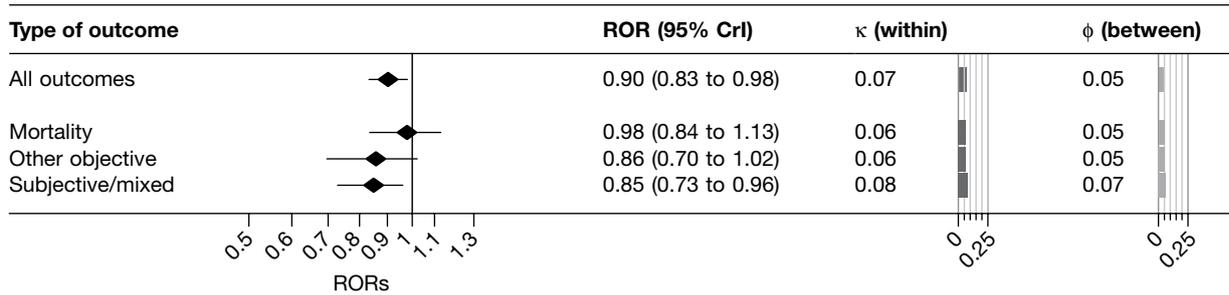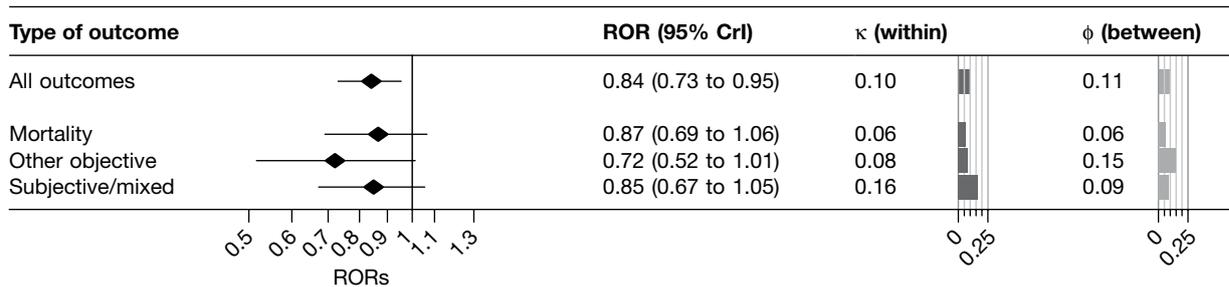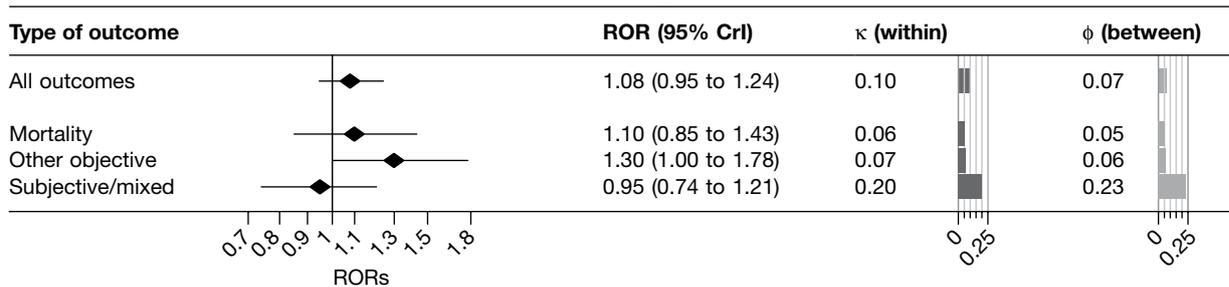
Inadequate or unclear allocation concealment (vs adequate)

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	0.90 (0.83 to 0.98)	0.07	0.05
Mortality	0.98 (0.84 to 1.13)	0.06	0.05
Other objective	0.86 (0.70 to 1.02)	0.06	0.05
Subjective/mixed	0.85 (0.73 to 0.96)	0.08	0.07

RORs

Lack of double blinding or unclear double blinding (vs double blind)

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	0.84 (0.73 to 0.95)	0.10	0.11
Mortality	0.87 (0.69 to 1.06)	0.06	0.06
Other objective	0.72 (0.52 to 1.01)	0.08	0.15
Subjective/mixed	0.85 (0.67 to 1.05)	0.16	0.09

RORs

Interaction between allocation concealment and double blinding

Type of outcome	ROR (95% CrI)	κ (within)	φ (between)
All outcomes	1.08 (0.95 to 1.24)	0.10	0.07
Mortality	1.10 (0.85 to 1.43)	0.06	0.05
Other objective	1.30 (1.00 to 1.78)	0.07	0.06
Subjective/mixed	0.95 (0.74 to 1.21)	0.20	0.23

RORs

FIGURE 6 Estimated RORs and effects on heterogeneity from multivariable analyses including interaction terms, according to type of outcome measure. Analyses based on trials with assessments of both allocation concealment and blinding. Upper panel, influence of inadequate or unclear versus adequate allocation concealment, in double-blind trials; middle panel, influence of lack of double blinding or unclear blinding status versus double blinding, in trials with adequate allocation concealment; lower panel, interaction between allocation concealment and blinding.

in heterogeneity was somewhat lower for lack of or unclear (compared with adequate) double blinding (0.24 95% CrI 0.02 to 0.79, based on seven meta-analyses containing 58 trials).

Results after excluding meta-analyses that contributed to the study of Wood et al.[16]

Data from the three contributing meta-epidemiological studies by Schulz et al.,[11] Kjaergard et al.[15] and Egger et al.[9] were combined in a study previously reported by Wood et al.,[16] which used a meta-meta-analytic approach for statistical analyses.[8] A total of 123 meta-analyses (1066 trials) were contributed by other meta-epidemiological studies. *Table 22* shows numbers of contributing meta-analyses and results from univariable analyses of the influence of the three study design characteristics, both overall and separately according to type of outcome measure. Compared with the full data set, estimated RORs tend to be closer to the null, except for the influence of inadequate or unclear (compared with adequate) sequence generation in meta-analyses with subjective outcomes (ROR 0.86, 95% CrI 0.75 to 0.97). However, effects on the between-trial heterogeneity are broadly consistent with those for the main analyses reported in *Table 14*, with between-trial heterogeneity increased in meta-analyses with subjective outcomes, for all study design characteristics.

TABLE 19 Estimated RORs and effects on heterogeneity associated with reported study design characteristics: results from multivariable analyses[a]

Study design characteristic and outcome	ROR	95% CrI	κ	95% CrI	φ	95% CrI
Inadequate or unclear sequence generation (vs adequate)						
All	0.90	0.82 to 0.99	0.06	0.01 to 0.20	0.05	0.01 to 0.15
Mortality	0.86	0.69 to 1.06	0.08	0.01 to 0.31	0.06	0.01 to 0.28
Objective	1.00	0.84 to 1.20	0.07	0.01 to 0.30	0.07	0.01 to 0.27
Subjective	0.88	0.76 to 1.00	0.05	0.01 to 0.21	0.06	0.01 to 0.24
Inadequate or unclear allocation concealment (vs adequate)						
All	0.89	0.81 to 0.99	0.06	0.01 to 0.19	0.05	0.01 to 0.18
Mortality	1.03	0.82 to 1.31	0.07	0.01 to 0.30	0.07	0.01 to 0.33
Objective	0.92	0.76 to 1.12	0.06	0.01 to 0.24	0.06	0.01 to 0.29
Subjective	0.82	0.70 to 0.94	0.08	0.01 to 0.27	0.07	0.01 to 0.30
Lack of double blinding or unclear double blinding (vs double blind)						
All	0.86	0.73 to 0.98	0.20	0.02 to 0.39	0.17	0.03 to 0.32
Mortality	1.07	0.78 to 1.48	0.09	0.01 to 0.44	0.08	0.01 to 0.42
Objective	0.91	0.64 to 1.33	0.10	0.01 to 0.50	0.20	0.02 to 0.85
Subjective	0.77	0.61 to 0.93	0.24	0.02 to 0.45	0.20	0.04 to 0.39

a   Numbers of contributing meta-analyses and trials are the same as in *Table 16*.

Univariable three-category analyses

Table 23 presents results from analyses of all trials in which reported methodological characteristics were assessed in three categories (low, unclear or high risk of bias). The number of trials and meta-analyses contributing to each analysis are shown in *Table 24*. There were no consistent patterns when comparing RORs for unclear (compared with low) and high (compared with low) risk of bias, which provides some support for combining these effects in the main analyses. Consistent with the main analyses, estimated values of κ were greatest for meta-analyses with subjectively assessed outcomes, both for high and unclear (compared with low) risk of bias.

Downweighting potentially biased evidence in future meta-analyses

Table 25 presents implications of the results from the primary univariable analyses (see *Table 14*) for downweighting of potentially biased evidence in future meta-analyses, based on formulae from Welton *et al.*[26] Because estimated values of κ and φ were greatest for meta-analyses with subjectively assessed outcomes, the minimum variance of the estimated intervention effect for a trial at high or unclear risk of bias is greatest for such trials. Across all BRANDO trials with inadequate or unclear sequence generation, bias adjustment led to a median 10% (IQR 4% to 23%) increase in trial-level variance. Downweighting based on results specific to type of outcome measure has the greatest effect in trials with subjectively assessed outcomes [median 20% (IQR 8% to 39%) increase in variance]. Results were broadly similar for downweighting based on inadequate or unclear allocation concealment. The median increase in variance for trials with subjectively measured outcomes that were not double blind or unclearly blinded was 63% (IQR 22% to 138%).

Downweighting all trials with inadequate or unclear sequence generation led to a median 13% (IQR 5% to 32%) increase in the variance of the summary (meta-analytic) intervention effect estimate among informative meta-analyses in the BRANDO database. This is in contrast to a median increase of 217% (IQR 87% to 482%) that results from completely excluding such trials,
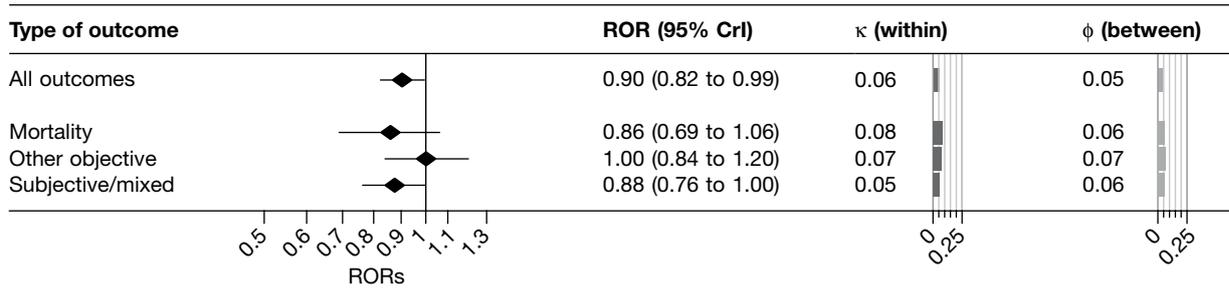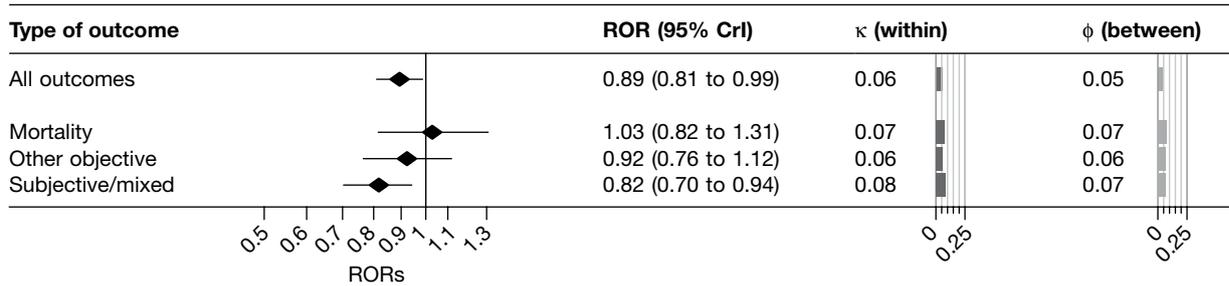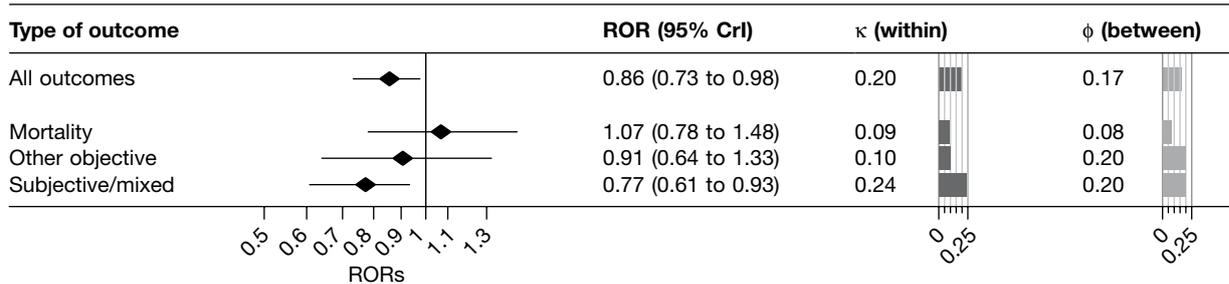
Inadequate or unclear generation of randomisation sequence (vs adequate)

Type of outcome		ROR (95% CrI)	κ (within)		φ (between)	
All outcomes		0.90 (0.82 to 0.99)	0.06		0.05	
Mortality		0.86 (0.69 to 1.06)	0.08		0.06	
Other objective		1.00 (0.84 to 1.20)	0.07		0.07	
Subjective/mixed		0.88 (0.76 to 1.00)	0.05		0.06	

RORs

Inadequate or unclear allocation concealment (vs adequate)

Type of outcome		ROR (95% CrI)	κ (within)		φ (between)	
All outcomes		0.89 (0.81 to 0.99)	0.06		0.05	
Mortality		1.03 (0.82 to 1.31)	0.07		0.07	
Other objective		0.92 (0.76 to 1.12)	0.06		0.06	
Subjective/mixed		0.82 (0.70 to 0.94)	0.08		0.07	

RORs

Lack of double blinding or unclear double blinding (vs double blind)

Type of outcome		ROR (95% CrI)	κ (within)		φ (between)	
All outcomes		0.86 (0.73 to 0.98)	0.20		0.17	
Mortality		1.07 (0.78 to 1.48)	0.09		0.08	
Other objective		0.91 (0.64 to 1.33)	0.10		0.20	
Subjective/mixed		0.77 (0.61 to 0.93)	0.24		0.20	

RORs

FIGURE 7 Estimated RORs and effects on heterogeneity from multivariable analyses including all three study design characteristics, according to type of outcome measure. Analyses based on trials with assessments of all three study design characteristics. Upper panel, influence of inadequate or unclear versus adequate sequence generation, adjusted for allocation concealment and blinding; middle panel, influence of inadequate or unclear versus adequate allocation concealment, adjusted for sequence generation and blinding; lower panel, influence of lack of double blinding or unclear blinding status versus double blinding, adjusted for sequence generation and allocation concealment.

because only 26% of trials were assessed to have adequate sequence generation. Bias adjustment for meta-analyses with subjectively assessed outcomes led to a median 31% (IQR 11% to 56%) increase in the variance of the summary intervention effect estimate, which was again small compared with complete exclusion of such trials. Results were broadly similar for the other study design characteristics, although differences between the effects of downweighting and excluding trials at high or unclear risk of bias were smaller for double blinding, because 56% of trials from informative meta-analyses were double blind. Even for subjectively assessed outcomes, excluding trials not assessed as double blind led to a greater loss of precision than retaining but downweighting them [median increase in variance 36% (IQR 8% to 72%) for downweighting compared with median 62% (IQR 19% to 140%) for excluding].

TABLE 20 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, stratified by clinical area: univariable analyses based on available data for each characteristic[a]

Study design characteristic by clinical area	No. of meta-analyses	No. of trials	Contributing meta-analyses[b]	Contributing trials	No. (%) of trials at high risk of bias[c]	ROR (95% CrI)	κ (95% CrI)	φ (95% CrI)
Pregnancy and childbirth								
Sequence generation	53	298	31 (15)	214	148 (69.2)	0.85 (0.69 to 1.05)	0.09 (0.01 to 0.33)	0.10 (0.01 to 0.47)
Allocation concealment	57	407	44 (26)	325	209 (64.3)	0.83 (0.70 to 0.97)	0.17 (0.02 to 0.47)	0.09 (0.01 to 0.33)
Blinding	57	447	29 (19)	249	129 (51.8)	0.78 (0.63 to 0.97)	0.29 (0.02 to 0.63)	0.10 (0.01 to 0.42)
Mental and behavioural								
Sequence generation	21	207	15 (4)	171	138 (80.7)	0.82 (0.67 to 1.01)	0.10 (0.01 to 0.29)	0.09 (0.01 to 0.42)
Allocation concealment	25	287	8 (5)	144	119 (82.6)	1.01 (0.77 to 1.40)	0.20 (0.02 to 0.36)	0.14 (0.01 to 0.58)
Blinding	26	302	6 (4)	145	42 (29.0)	0.91 (0.43 to 1.35)	0.14 (0.01 to 0.41)	0.27 (0.02 to 1.17)
Circulatory system								
Sequence generation	18	119	11 (8)	100	72 (72.0)	0.90 (0.72 to 1.08)	0.13 (0.01 to 0.39)	0.06 (0.01 to 0.33)
Allocation concealment	31	257	24 (13)	216	148 (68.5)	0.98 (0.87 to 1.09)	0.06 (0.01 to 0.19)	0.04 (0.01 to 0.17)
Blinding	31	277	20 (12)	187	83 (44.4)	0.92 (0.79 to 1.04)	0.06 (0.01 to 0.20)	0.06 (0.01 to 0.24)

a   For each trial characteristic, comparisons are the same as in previous tables.
b   Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
c   High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.

TABLE 21 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, stratified by type of intervention: univariable analyses based on available data for each characteristic[a]

Bias domain by intervention type	No. of meta-analyses	No. of trials	Contributing meta-analyses[b]	Contributing trials	No. (%) of trials at high risk of bias[c]	ROR (95% CrI)	κ (95% CrI)	φ (95% CrI)
Pharmacological interventions								
Sequence generation	124	869	76 (38)	691	519 (75.1)	0.86 (0.78 to 0.95)	0.16 (0.03 to 0.27)	0.05 (0.01 to 0.18)
Allocation concealment	157	1301	108 (63)	1012	713 (70.5)	0.88 (0.81 to 0.94)	0.12 (0.02 to 0.23)	0.04 (0.01 to 0.14)
Blinding	162	1418	94 (59)	987	420 (42.6)	0.87 (0.77 to 0.96)	0.13 (0.02 to 0.29)	0.16 (0.03 to 0.31)
Surgical interventions								
Sequence generation	11	58	6 (5)	48	30 (62.5)	0.93 (0.47 to 1.79)	0.29 (0.02 to 1.36)	0.11 (0.01 to 0.92)
Allocation concealment	14	107	9 (5)	66	50 (75.8)	1.41 (0.52 to 2.58)	0.15 (0.01 to 0.88)	0.15 (0.01 to 1.94)
Blinding	No data							

a   For each study design characteristic, comparisons are the same as in previous tables.
b   Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
c   High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.

TABLE 22 Estimated RORs and effects on heterogeneity associated with reported study design characteristics in a data set excluding meta-analyses used in the study of Wood *et al*.:[16] univariable analyses based on available data for each characteristic

Study design characteristic and outcome	No. of meta-analyses	No. of trials	Contributing meta-analyses[a]	Contributing trials	No. (%) of trials at high risk of bias[b]	ROR (95% CrI)	κ (95% CrI)	φ (95% CrI)
Inadequate or unclear sequence generation (vs adequate)								
All	103	837	74 (40)	690	516 (74.8)	0.91 (0.83 to 1.00)	0.16 (0.02 to 0.29)	0.05 (0.01 to 0.17)
Mortality	17	100	11 (7)	76	57 (75.0)	0.94 (0.71 to 1.23)	0.07 (0.01 to 0.37)	0.08 (0.01 to 0.43)
Objective	31	253	26 (12)	221	159 (71.9)	1.03 (0.84 to 1.26)	0.13 (0.01 to 0.54)	0.07 (0.01 to 0.29)
Subjective	55	484	37 (21)	393	300 (76.3)	0.86 (0.75 to 0.97)	0.19 (0.02 to 0.32)	0.06 (0.01 to 0.21)
Inadequate or unclear allocation concealment (vs adequate)								
All	123	1066	68 (39)	727	551 (75.8)	0.98 (0.90 to 1.07)	0.09 (0.01 to 0.22)	0.04 (0.01 to 0.15)
Mortality	25	191	16 (8)	147	105 (71.4)	1.00 (0.86 to 1.14)	0.07 (0.01 to 0.24)	0.05 (0.01 to 0.22)
Objective	32	261	18 (11)	173	124 (71.7)	0.96 (0.78 to 1.20)	0.07 (0.01 to 0.30)	0.08 (0.01 to 0.41)
Subjective	66	614	34 (20)	407	322 (79.1)	0.97 (0.85 to 1.11)	0.16 (0.02 to 0.32)	0.07 (0.01 to 0.25)
Lack of double blinding or unclear double blinding (vs double blind)								
All	123	1063	45 (28)	525	237 (45.1)	0.91 (0.80 to 1.02)	0.09 (0.01 to 0.28)	0.14 (0.02 to 0.32)
Mortality	25	191	12 (7)	115	55 (47.8)	0.89 (0.74 to 1.06)	0.06 (0.01 to 0.22)	0.08 (0.01 to 0.32)
Objective	32	260	8 (5)	109	47 (43.1)	0.90 (0.50 to 1.92)	0.16 (0.01 to 0.87)	0.38 (0.02 to 1.48)
Subjective	66	612	25 (16)	301	135 (44.9)	0.93 (0.74 to 1.10)	0.22 (0.02 to 0.42)	0.14 (0.02 to 0.45)

a Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).
b High risk of bias means inadequate/unclear random sequence generation or allocation concealment or unclear/not double blind.

TABLE 23 Estimated RORs and effects on heterogeneity associated with reported study design characteristics, according to the type of outcome measure: univariable analyses based on all trials with study design characteristics assessed in three categories

Study design characteristic and outcome	No. of meta-analyses	No. of trials	ROR (95% CrI)		κ (95% CrI)		φ (95% CrI)	
			Unclear[a]	High[a]	Unclear[a]	High[a]	Unclear[a]	High[a]
Sequence generation								
All	186	1176	0.90 (0.83 to 0.98)	0.78 (0.62 to 0.98)	0.06 (0.01 to 0.16)	0.49 (0.11 to 0.75)	0.05 (0.01 to 0.15)	0.07 (0.01 to 0.34)
Mortality	30	137	0.88 (0.68 to 1.11)	1.34 (0.70 to 2.84)	0.06 (0.01 to 0.20)	0.08 (0.01 to 0.50)	0.07 (0.01 to 0.34)	0.14 (0.01 to 1.41)
Objective	66	387	1.02 (0.85 to 1.21)	0.87 (0.57 to 1.29)	0.05 (0.01 to 0.18)	0.38 (0.02 to 1.03)	0.07 (0.01 to 0.28)	0.11 (0.01 to 0.67)
Subjective	90	652	0.85 (0.75 to 0.96)	0.66 (0.46 to 0.90)	0.11 (0.01 to 0.25)	0.57 (0.22 to 0.91)	0.06 (0.01 to 0.20)	0.09 (0.01 to 0.55)
Allocation concealment								
All	228	1796	0.92 (0.86 to 0.99)	0.95 (0.82 to 1.10)	0.13 (0.02 to 0.22)	0.21 (0.02 to 0.48)	0.04 (0.01 to 0.13)	0.09 (0.01 to 0.32)
Mortality	44	328	0.98 (0.87 to 1.10)	1.03 (0.69 to 1.63)	0.07 (0.01 to 0.19)	0.19 (0.01 to 1.06)	0.05 (0.01 to 0.17)	0.15 (0.01 to 0.70)
Objective	76	551	0.96 (0.84 to 1.10)	1.12 (0.78 to 1.65)	0.06 (0.01 to 0.20)	0.18 (0.01 to 0.86)	0.06 (0.01 to 0.25)	0.17 (0.01 to 0.68)
Subjective	108	917	0.84 (0.76 to 0.94)	0.86 (0.71 to 1.03)	0.22 (0.08 to 0.32)	0.18 (0.02 to 0.50)	0.06 (0.01 to 0.22)	0.08 (0.01 to 0.35)
Blinding								
All	224	1721	0.90 (0.69 to 1.15)	0.82 (0.70 to 0.96)	0.14 (0.01 to 0.44)	0.11 (0.01 to 0.23)	0.16 (0.01 to 0.59)	0.23 (0.06 to 0.42)
Mortality	44	304	0.86 (0.39 to 1.57)	0.94 (0.69 to 1.26)	0.15 (0.01 to 1.04)	0.06 (0.01 to 0.22)	0.19 (0.01 to 1.33)	0.08 (0.01 to 0.41)
Objective	77	591	1.30 (0.76 to 2.22)	0.89 (0.67 to 1.25)	0.08 (0.01 to 0.35)	0.05 (0.01 to 0.18)	0.11 (0.01 to 0.74)	0.16 (0.01 to 0.65)
Subjective	103	826	0.75 (0.51 to 1.09)	0.70 (0.53 to 0.89)	0.26 (0.02 to 0.77)	0.27 (0.08 to 0.38)	0.15 (0.01 to 0.70)	0.33 (0.13 to 0.59)

a   Compared with low risk of bias (adequate sequence generation or allocation concealment, or double blind).

TABLE 24 The number of trials and meta-analyses contributing to the three-category analyses presented in *Table 23*

Study design characteristic and outcome	High vs low risk of bias			Unclear vs low risk of bias		
	Contributing meta-analyses[a]	Contributing trials	Contributing high-risk trials (%)	Contributing meta-analyses[a]	Contributing trials	Contributing unclear-risk trials (%)
Sequence generation						
All	48 (17)	453	81 (18)	100 (57)	906	648 (72)
Mortality	3 (2)	26	7 (27)	12 (8)	83	59 (71)
Objective	22 (6)	175	32 (18)	38 (18)	303	204 (67)
Subjective	23 (9)	252	42 (17)	50 (31)	520	385 (74)
Allocation concealment						
All	75 (29)	716	136 (19)	171 (99)	1485	998 (67)
Mortality	11 (2)	95	17 (18)	35 (18)	287	187 (65)
Objective	26 (6)	215	38 (18)	54 (30)	442	293 (66)
Subjective	38 (21)	406	81 (20)	82 (51)	756	518 (69)
Blinding						
All	71 (36)	725	290 (40)	33 (17)	286	84 (29)
Mortality	13 (5)	119	39 (33)	5 (3)	48	10 (21)
Objective	20 (12)	208	81 (39)	9 (5)	63	23 (37)
Subjective	38 (19)	398	170 (43)	19 (9)	175	51 (29)

a   Numbers in parentheses are meta-analyses that contribute to estimation of κ (at least two trials with and without characteristic).

TABLE 25 Estimated downweighting (increase in variance) in trials and meta-analyses included in the BRANDO data set corresponding to the estimated effects of reported study design characteristics shown in *Table 14*, for each characteristic and type of outcome measure

Study design characteristic and outcome	No. of high-risk trials	Minimum variance of trial at high risk of bias ($V_0 + \kappa^2 + \varphi^2$)	Median (IQR) increase in trial-level variance (%)	Median (IQR) increase in variance of summary intervention effect (%)		
				Downweighting all meta-analyses	Downweighting informative meta-analyses	Excluding all trials at high or unclear risk of bias
Inadequate or unclear sequence generation (vs adequate)						
All	901	0.030	10 (4 to 23)	12 (2 to 32)	13 (5 to 32)	217 (87 to 482)
Mortality	116	0.020	6 (3 to 14)	11 (1 to 25)	13 (6 to 36)	119 (70 to 336)
Objective	273	0.019	5 (3 to 11)	8 (1 to 19)	11 (2 to 32)	145 (62 to 559)
Subjective	512	0.046	20 (8 to 39)	31 (6 to 64)	31 (11 to 56)	282 (126 to 482)
Inadequate or unclear allocation concealment (vs adequate)						
All	1380	0.017	5 (2 to 12)	9 (3 to 23)	7 (3 to 20)	150 (49 to 411)
Mortality	233	0.011	4 (1 to 11)	8 (3 to 34)	8 (3 to 19)	121 (39 to 468)
Objective	413	0.011	3 (1 to 6)	9 (3 to 22)	6 (3 to 13)	175 (52 to 337)
Subjective	734	0.053	18 (7 to 40)	36 (8 to 73)	27 (7 to 59)	146 (55 to 411)
Lack of double blinding or unclear double blinding (vs double blind)						
All	1041	0.044	13 (6 to 31)	16 (0 to 62)	15 (3 to 48)	62 (19 to 143)
Mortality	170	0.013	4 (2 to 10)	5 (0 to 18)	5 (0 to 18)	46 (18 to 101)
Objective	336	0.036	11 (5 to 24)	22 (0 to 67)	16 (3 to 46)	79 (22 to 202)
Subjective	535	0.200	63 (22 to 138)	41 (1 to 175)	36 (8 to 72)	62 (19 to 140)

Chapter 4

Discussion

Summary of findings

Using data from 10 empirical (meta-epidemiological) studies, we developed a combined database for meta-epidemiological research. The database structure comprised six tables with defined relationships between them, which reflects the complexity of relationships between reviews, their publications, meta-analyses, trials and their publications, and trial characteristics and results. This database structure allowed us to identify duplicated entries, so that in the final database there were no overlaps between meta-analyses. This database design is potentially relevant to any situation in which collections of meta-analyses are being combined, and may also be of interest to researchers wishing to identify overlaps between two or more previously published meta-analyses.

We estimated the influence of three types of study design characteristic on average intervention effect estimates, and in increasing between-trial, within-meta-analysis, heterogeneity, in 1973 trials (234 meta-analyses) for which information on both study design characteristics and trial results was available. Bias in intervention effect estimates resulting from inadequate or unclear sequence generation, inadequate or unclear allocation concealment or lack of or unclear double blinding varied according to the type of outcome measure assessed. Overall, there was little evidence of bias in trials assessing all-cause mortality or other objectively assessed outcomes. In contrast, inadequate sequence generation, inadequate allocation concealment or lack of blinding were associated with exaggerated estimates of the benefit of interventions in trials reporting subjectively assessed outcomes. The direction and magnitude of bias associated with reported study design characteristics varies between trials and meta-analyses, and with the type of outcome measure. Increases in between-trial heterogeneity associated with study design characteristics were greatest for trials reporting subjectively assessed outcomes. Except for the effect of lack of or unclear double blinding in trials reporting subjectively assessed outcomes, estimates of between-meta-analysis variability in mean bias were small. Analyses of the effects of combined study design characteristics suggested that effects of individual characteristics were less than multiplicative, in that estimated effects of two study design characteristics together were attenuated compared with the combined individual effects. Effects were somewhat attenuated in multivariable analyses.

Strengths and weaknesses

To our knowledge, this study represents the most comprehensive attempt to date to quantify the influence of inadequate or unclear random sequence generation, inadequate or unclear allocation concealment and lack of or unclear double blinding on intervention effect estimates from RCTs. So far as we are aware, no previous study has quantified the effect of reported study design characteristics on between-trial heterogeneity as well as on average intervention effects, although Schulz et al.[11] noted that ORs from inadequately concealed trials appeared more heterogeneous than those from adequately concealed trials. Thanks to the generosity of their investigators, we were able to combine the results of all meta-epidemiological studies of which we were aware at the start of the study, with the exception of a study for which data were no longer available.

The size of the data set meant that we were able to quantify the influence of study design characteristics in subgroups defined by types of intervention, comparison and outcome measure.

Combining multiple collections of meta-analyses and removing overlaps is a labour-intensive process. We estimate that for the very large BRANDO study database this process took approximately 24 months' person time. However, a substantial proportion of this time was devoted to database development and programming, which might be avoided in future studies. Harmonising the definition of trial characteristics across different studies was also time-consuming. Nonetheless, de novo data extraction would have taken substantially longer. It would be difficult to publish results from a combined database in which the extent of overlap between meta-analyses was unknown. However, combining meta-epidemiological studies on different topics would likely avoid the need for a deduplication exercise.

Inadequate reporting of trial methods can severely impede the assessment of trial quality and the risk of bias in trial results. This is a particular problem for the assessment of sequence generation and allocation concealment, which are often not described at all in trial publications. *Table 11* shows that, for 64% of trials with assessment of sequence generation and 69% of trials with assessments of allocation concealment, these characteristics were assessed as unclear. In most cases this was because the method was not described at all in the publication. Transparent and standardised reporting of trials is an essential part of clinical research. Many trials included in this data set were published before the publication of the CONSORT (Consolidated Standards of Reporting Trials) recommendations in 1996.[31,32] There is evidence that reporting of trials improved between 2000 and 2006, although it remained below an acceptable level.[33]

Our study was based on reported study design characteristics, which need not correspond to how a trial was in fact conducted: well-conducted trials may be reported badly.[34] A study of discrepancies between published reports and the actual conduct of RCTs found that, in more than three-quarters of trials in which sequence generation or allocation concealment were unclear, these characteristics were in fact adequate,[35] whereas Soares *et al.*[36] found that the methodology of 56 trials of radiation therapy in oncology was better than reported. In contrast, Pildal *et al.*[37] found that most trials with unclear allocation concealment on the basis of the trial publication also have unclear allocation concealment according to their protocol. A recent study of reporting of blinding suggested that, although reporting is often inadequate, it rarely contradicts the methods specified in the trial protocol.[38]

The study of Egger *et al.*[9] relied on assessments of trial quality by the authors of the included Cochrane reviews rather than by methodological experts. Despite the standardised guidelines specified in the *Cochrane Handbook*[39] at the time, evaluations by authors of Cochrane reviews of whether or not a study had adequate allocation concealment may be inconsistent.[37] The effect of trial quality on estimates of intervention effect in this study was, however, in line with previous studies in which quality was assessed by the same observers: one would expect attenuation of effects if assessments in Cochrane reviews were less reliable.

As is common in Bayesian hierarchical modeling,[28,29] we found a high degree of sensitivity of estimated variance components to the prior distributions assumed for these parameters. This is of concern because the estimates of κ and φ are of substantial interest, having the potential to drive downweighting of evidence from high or unclear risk trials in future meta-analyses, based on formulae from Welton *et al.*[26] Lambert *et al.*[28] have previously demonstrated an upward bias of Monte Carlo Markov Chain estimates of variance components when the true variance is very close to zero. We were, therefore, cautious in interpreting small estimates of heterogeneity parameters, and those for which the lower limit of the CI is close to zero.

We estimated the impact of downweighting trials in a new fixed-effect meta-analysis, based on our estimates of the influence of study design characteristics on intervention effect estimates. In practice, a random-effects model might have been preferable for many of the meta-analyses on which these results are based. The hierarchical bias model used to estimate κ^2, φ^2 and V_0 assumed random intervention effects within each meta-analysis, and can be expected to produce estimates of κ^2 and φ^2 that are smaller than those estimated from a fixed-treatment-effect model. These results should therefore be interpreted only as an approximate guide to the likely impact of bias adjustment in a new fixed-effect meta-analysis.

The present study in context with other literature

Because different meta-epidemiological studies have reported on similar associations (e.g. the association of inadequate or unclear concealment of allocation with intervention effect estimates), some authors have previously reported meta-analytic estimates combined across different meta-epidemiological studies.[14] The validity of such estimates would be undermined if different meta-epidemiological studies were reporting on the same meta-analyses, but the extent of overlap has until now been unclear. Our study provides some reassurance in that only 69 (21%) meta-analyses (492 trial results, 16%) were contained in more than one meta-epidemiological study that had both recorded trial results and assessed study design characteristics.

The overlap between different meta-epidemiological studies provided an opportunity to examine the reliability of assessments of sequence generation, allocation concealment and double blinding between pairs of contributing studies. Overall, we found good agreement. This is encouraging given our aim of conducting combined analyses, particularly as assessments were carried out completely independently, and because definitions were not completely consistent between the different contributing studies. Double blinding was the characteristic for which there was the highest inter-rater agreement; this may be because the presence of and the methods for blinding tend to be better reported than the methods of sequence generation and concealment. It is relatively easy to identify that a trial has been reported as either 'double blind' or 'open', whereas no such standard terms exist for the randomisation process. A recent study by Hartling et al.[40] assessed inter-rater agreement between two independent reviewers who applied the Cochrane Collaboration's risk of bias tool on a convenience sample of 163 published randomised trials. They found moderate to good agreement for sequence generation, allocation concealment and blinding, but in contrast to our study they observed the highest agreement for sequence generation (kappa = 0.74, 95% CI 0.64 to 0.85) and lowest for the assessment of blinding (kappa = 0.35; 95% CI 0.22 to 0.47). This could be because the new Cochrane guidelines[41] for risk of bias assessment state that it is inappropriate to pass a judgement of a low risk of bias if the study is merely described as 'double blind' without further details, thus rendering the assessment of blinding less straightforward. Accurate measurement of methodological characteristics is essential for the validity of meta-epidemiological studies. Given that the average effects of such characteristics are modest,[6,16] non-differential misclassification may dilute or even extinguish them, whereas differential misclassification could create spurious effects.

Despite the large number of meta-epidemiological studies that contributed to the BRANDO study, only the three study design characteristics, sequence generation, allocation concealment and double blinding, were consistently assessed across these studies. The Cochrane Collaboration's risk of bias tool, which was first published in 2008[41] and was recently updated,[42] includes assessment of the risk of bias due to incomplete outcome data and selective reporting of outcomes. There is recent empirical evidence of bias in the results of RCTs as a result of both attrition[43,44] and selective outcome reporting.[2,45,46] It is possible that our results were confounded by the influence of these or other types of bias, which could not be adjusted for. However,

estimated effects of the study design characteristics analysed in our study were only modestly attenuated in adjusted analysis. The influence of study design characteristics may also vary in clinical areas that were not well represented in the BRANDO database.[47]

Definitions of adequate sequence generation and allocation concealment were relatively consistent between the contributing studies (see *Table 9*). For allocation concealment, all studies adhered to the definitions originally formulated by Schulz *et al.*[11] In contrast, definitions used to assess a trial as double blind varied somewhat (see *Table 9*), with some being more strict[11,14] than others: three studies[9,12,21] considered a trial double blind merely if the trial report described it as double blind. However, we found that the assessment of double blinding was consistent between the contributing studies, with a median κ statistic of 0.87.[17] The term 'double blind' is problematic because at least three distinct groups (trial participants, trial personnel and outcome assessors) can potentially be blinded. Both physicians and textbooks vary in their interpretations and definitions of 'single', 'double' and 'triple' blinding.[48] We hope that future trials will report the blinding status of the different groups involved, as specified in the CONSORT statement,[31,32] rather than using the vague term 'double blind'.[49] The recent update of the Cochrane Collaboration's risk of bias tool[42] separates assessment of the risk of performance bias (adequacy of blinding of participants and personnel) from assessment of the risk of detection bias (adequacy of blinding of outcome assessors).

Implications for research

The database structure that we developed may have wide applicability in evidence synthesis research. It has the potential to ensure data integrity and consistency in situations in which it is necessary to store data from multiple systematic reviews, or to accommodate multiple publications and multiple study designs in large systematic reviews. The database structure can be modified to store information on a variety of study types and additional facilities for duplicate data entry and checking have also been developed for different projects using the basic structure.

We found that a lack of or unclear double blinding was associated with marked exaggeration of intervention effect estimates and increases in between-trial heterogeneity in trials with subjectively assessed outcome measures. This is consistent with there being more prominent placebo effects in such trials.[50,51] It is possible that placebo effects are themselves of clinical utility. It would be of great interest to examine separate effects of blinding of participants and personnel (performance bias) and blinding of outcome assessors (detection bias) once such assessments are available from large collections of trials and meta-analyses.

In contrast to previous studies,[9,11] we found the influence of lack of double blinding to be greater than that of inadequate or unclear random sequence generation or allocation concealment. Our finding that the influence of these last two study design characteristics is most marked for trials with subjectively assessed outcomes was unexpected, although it is consistent with the results of Wood *et al.*[16] for allocation concealment. The purpose of random sequence generation and allocation concealment is to avoid selection bias, whereby knowledge of prognosis at the time of recruitment to a randomised trial influences the intervention group to which the patient is allocated.[52] Such selection bias would be expected to be greatest when it is easy to assess patients' prognosis at the time they are recruited to a trial, and affect the results of trials with objectively assessed as well as subjectively assessed outcome measures. As shown in *Table 12*, presence of double blinding was associated with adequate allocation concealment but not with adequate sequence generation. The influence of sequence generation and allocation concealment in trials with subjectively assessed outcomes was little attenuated in multivariable analyses. Therefore, these effects may result at least in part from their association with subsequent flaws in the

conduct of trials, in particular with biased outcome assessment, rather than from selection bias. Adequate sequence generation and allocation concealment may also be markers for other strategies for reducing bias, beyond blinding.

Ownership of the BRANDO database remains with the investigators of the contributing studies. However, the steering group would be happy to consider requests for access to the database once the main study results are published.

Recommendations for future research

Tools for assessing risk of bias in results of RCTs[41,42] and guidelines for summarising the quality of evidence from systematic reviews[53] should account for the findings of this study. In particular, it appears that when study design characteristics associated with bias prevention are not in place, trial results based on subjectively assessed outcome measures are at greatest risk of bias.

Practical and acceptable methods for correcting and downweighting the results of trials at high risk of bias in new meta-analyses should be developed. The trade-off between bias and variation is a familiar problem in statistics: our results suggest that exclusion of trials at high risk of bias leads to much greater decreases in precision than empirically based downweighting of such studies.

The influence of further study design characteristics should be explored in new meta-epidemiological studies based on more recently reported trials than are available in the BRANDO database. These include the influence of incomplete outcome data and selective reporting of outcomes, examining separate effects of blinding of participants and personnel (performance bias) and blinding of outcome assessors (detection bias) and the influence of characteristics specific to particular study designs (e.g. carry-over effects in crossover trials). It would also be of interest to examine clinical areas that are not well represented in the BRANDO data set.

Bayesian models could be extended to deal with misclassification of study design characteristics, based on empirical evidence such as that presented in *Table 5*.

Our results suggest that, as far as possible, clinical and policy decisions should not be based on trials in which blinding is not feasible and outcome measures are subjectively assessed. Therefore, trials in which blinding is not feasible should focus as far as possible on objectively measured outcomes, and should aim to blind outcome assessors.

Acknowledgements

Contribution of authors

Dr Jelena Savović (Postdoctoral Fellow, Epidemiology and Evidence Synthesis) contributed to the study design, designed the database, combined and harmonised data contributed by co-authors, removed overlapping trials and meta-analyses from the database, extracted additional data from reviews, carried out most of the data cleaning and data management, co-ordinated the study and contacts with all co-authors, carried out descriptive analyses, co-wrote the first draft and contributed to revisions of the manuscript.

Dr Hayley E Jones (Research Fellow, Medical Statistics, Bayesian Modelling and Evidence Synthesis) conducted most analyses, wrote sections of the first draft and contributed to the revisions of the manuscript.

Professor Douglas G Altman (Professor, Statistics in Medicine) played a key role in the design and planning, and in making strategic decisions throughout the project, and contributed to the database design and to harmonising and integration of the contributed data, and to the drafting and editing of the manuscript.

Ross J Harris (Statistician, Medical Statistics, Bayesian Modelling and Evidence Synthesis) carried out data manipulation and identification of overlapping trials, conducted initial analyses, refined the statistical model and contributed to the editing of the manuscript.

Professor Peter Jűni (Professor and Head of Division, Clinical Epidemiology and Biostatistics) provided raw data from his original study, advised on the interpretation and integration of data from his study into the combined database and on extraction and classification of additional data, and contributed to the steering of the project and to the editing of the manuscript.

Dr Julie Pildal (Specialist Registrar, Endocrinology and Research Synthesis) provided raw data from her original study, advised on the interpretation and integration of data from her study into the combined database and on extraction and classification of additional data, and contributed to the steering of the project and to the editing of the manuscript.

Dr Bodil Als-Nielsen (Specialist Registrar, Paediatrics) provided raw data from her original study, advised on the interpretation and integration of data from her study into the combined database and on extraction and classification of additional data, and contributed to the steering of the project and to the editing of the manuscript.

Dr Ethan M Balk (Associate Director, Clinical Evidence Synthesis) provided raw data from his original study, advised on the interpretation and integration of data from his study into the combined database and on extraction and classification of additional data, and contributed to the steering of the project and to the editing of the manuscript.

Publication

1. Savović J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, *et al.* Development of a combined database for meta-epidemiological research. *Res Syn Meth* 2010;**1**:212–25.

2. Savovic J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, *et al.* Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;**157**:429–38.

References

1. Altman DG, Schulz KF. Statistics notes: concealing treatment allocation in randomised trials. *BMJ* 2001;**323**:446–7.

2. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *J Am Med Assoc* 2004;**291**:2457–65.

3. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *J Am Med Assoc* 1992;**267**:374–8.

4. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;**1**:MR000006.

5. Scherer RW, Langenberg P, von EE. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;**2**:MR000005.

6. Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;**163**:493–501.

7. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997;**315**:617–19.

8. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21**:1513–24.

9. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;**7**(1).

10. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.

11. 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. *J Am Med Assoc* 1995;**273**:408–12.

12. Als-Nielsen B, Chen W, Gluud LL, Siersma V, Hilden J, Gluud C. Are trial size and reported methodological quality associated with treatment effects? Observational study of 523 randomised trials. 12th Cochrane Colloquium: Bridging the Gaps, Ottawa, ON, Canada, 2–6 October 2004.

13. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, *et al.* Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *J Am Med Assoc* 2002;**287**:2973–82.

14. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;**36**:847–57.

15. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;**135**:982–9.

16. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, *et al.* Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;**336**:601–5.

17. Savovic J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, *et al.* Development of a combined database for meta-epidemiological research. *Res Syn Meth* 2010;**1**:212–25.

18. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion on grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;**356**:1228–31.

19. Royle P, Milne R. Literature searching for randomised controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care* 2003;**19**:1–13.

20. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, *et al.* Should meta-analysts search EMBASE in addition to MEDLINE? *J Clin Epidemiol* 2003;**56**:943–55.

21. Contopoulos-Ioannidis DG, Gilbody SM, Trikalinos TA, Churchill R, Wahlbeck K, Ioannidis JP. Comparison of large versus smaller randomized trials for mental health-related interventions. *Am J Psychiatry* 2005;**162**:578–84.

22. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007;**26**:2745–58.

23. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;**330**:1053.

24. Kennedy E, Song F, Hunter R, Clarke A, Gilbody S. Risperidone versus typical antipsychotic medication for schizophrenia. *Cochrane Database Syst Rev* 2000;**2**:CD000440.

25. Song F. Risperidone in the treatment of schizophrenia: a meta-analysis of randomized controlled trials. *J Psychopharmacol* 1997;**11**:65–71.

26. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc* 2009;**172**:119–36.

27. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;**10**:325–37.

28. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med* 2005;**24**:2401–28.

29. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006;**1**:515–33.

30. World Health Organization. *International statistical classification of diseases and related health problems 10th revision. Version for 2007*. URL: http://apps.who.int/classifications/apps/icd/icd10online/ (accessed 19 August 2010).

31. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, *et al. CONSORT* 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c869.

32. Schulz KF, Altman DG, Moher D. *CONSORT* 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c332.

33. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. *The quality of reports of randomised trials in* 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;**340**:c723.

34. Huwiler-Muntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *J Am Med Assoc* 2002;**287**:2801–4.

35. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;**55**:783–6.

36. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, *et al.* Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;**328**:22–4.

37. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;**330**:1049.

38. Hrobjartsson A, Pildal J, Chan AW, Haahr MT, Altman DG, Gotzsche PC. Reporting on blinding in trial protocols and corresponding publications was often inadequate but rarely contradictory. *J Clin Epidemiol* 2009;**62**:967–73.

39. Cochrane Reviewers' Handbook 3.0.2. In *The Cochrane Library. The Cochrane Collaboration*. Oxford: Update Software; 1997.

40. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs SJ, *et al.* Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;**339**:b4012.

41. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: John Wiley; 2008. pp. 187–241.

42. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928.

43. Nuesch E, Trelle S, Reichenbach S, Rutjes AWS, Burgi E, Scherer M, *et al.* The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 2009;**339**:b3244.

44. Vestbo J, Anderson JA, Calverley PMA, Celli B, Ferguson GT, Jenkins C, *et al.* Bias due to withdrawal in long-term randomised trials in COPD: evidence from the TORCH study. *Clin Respir J* 2011;**5**:44–9.

45. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;**3**:e3081.

46. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, *et al.* The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;**340**:c365.

47. Nuesch E, Reichenbach S, Trelle S, Rutjes AWS, Liewald K, Sterchi R, *et al.* The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum* 2009;**61**:1633–41.

48. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, *et al.* Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *J Am Med Assoc* 2001;**285**:2000–3.

49. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 2002;**55**:787–90.

50. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 2001;**344**:1594–602.

51. Krogsboll LT, Hrobjartsson A, Gotzsche PC. Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention. *BMC Med Res Methodol* 2009;**9**:1

52. Forder PM, Gebski VJ, Keech AC. Allocation concealment and blinding: when ignorance is bliss. *Med J Aust* 2005;**182**:87–9.

53. GRADE Working Group. URL: www.gradeworkinggroup.org (accessed 15 August 2011).

Appendix 1

Protocol

This HTA report describes the work conducted by the workstream A, as designated in the original funding proposal. The report of the work carried out by the workstream B will be published separately.

I.　SUMMARY OF PROPOSAL

TITLE OF PROJECT

Quantification of bias in randomised controlled trials and non-randomised studies: implications for systematic reviews and evidence synthesis

APPLICANTS (NOTE: Section IV should also be completed for ALL applicants)

A: LEAD APPLICANT

Surname(s):	Sterne
Forename(s):	Jonathan A C
Title:	Dr

B: OTHER APPLICANTS

List separately each individual involved in the research project, giving their name, title, and responsibility:

Name	Jon Deeks
Name:	Douglas Altman
Name:	Matthias Egger
Name:	David Moher
Name:	Lise Gluud
Name:	John Ioannidis
Name:	Kenneth F. Schulz
Name:	Pamela Royle
Name:	Julie Pildal

SUMMARY OF RESEARCH

ABSTRACT OF RESEARCH. No more than 200 words covering the following topics: aims of project; research subject group; sample size, type and location; methods of working.

This project aims to examine the importance of different types of bias and other sources of variability in estimates of the effect of health care interventions, in different settings and areas of medicine. An existing database will be extended to combine data from all existing 'meta-epidemiological' studies of the effect of methodological quality and other trial characteristics on intervention effect estimates in randomised controlled trials (RCTs). Based on the results of a review conducted by the applicants, we will also assemble a new database that will be used to compare results of RCTs and non-randomised studies (NRS) that estimated the effects of comparable interventions on comparable outcomes. We will use a database of NRS examining the association of diet and physical activity with cancer to quantify differences in the magnitude of associations between types of non-randomised study. Statistical methods developed by the applicants will be extended to quantify differences in intervention effect estimates and the between-meta-analysis variability in these differences. These will lead to further development of guidelines for the conduct and reporting of RCTs and NRS, and to evidence-based methods for combining results of RCTs of varying quality, and RCTs and NRS, in systematic reviews of health care interventions.

II   DETAILS OF PROPOSED RESEARCH

Background

Bias in randomised controlled trials

Randomised controlled trials (RCTs) provide the best evidence in the evaluation of medical interventions. However, the number of patients included in trials is often inadequate and single trials may fail to detect, or exclude with certainty, a modest but medically important difference in the effects of two therapies.[1,2] An examination of clinical trials which reported no statistically significant differences between experimental and control therapy showed that false negative results in health care research are common: for a medically important difference in outcome the probability of missing this effect given the trial size was greater than 20% in 115 (85%) of the 136 trials examined.[1] Similarly, a recent examination of 1941 trials relevant to the treatment of schizophrenia showed that only 58 (3%) studies were large enough to detect a modest but important improvement.[3] Although there is some evidence that the number of patients included in trials published in general health care journals has increased,[4] under-powered trials continue to be published in large numbers.[5]

Systematic reviews and meta-analyses aim to address this situation. Statistical combination of results from several small trials in meta-analysis will increase the precision of estimated treatment effect, reduce the probability of 'false negative' results, and potentially lead to more timely introduction of effective treatments.[6] Systematic reviews and meta-analyses provide the basis for development of sound clinical practice guidelines and, in general, for evidence based health care.[7,8]

Systematic reviews and meta-analyses are not immune to bias, however. The majority of published meta-analyses are based on relatively few, small trials.[9,10] Small trials tend to show bigger treatment effects than larger trials, a difference that may be due to publication bias and related reporting biases such as time lag bias or language bias, or bias due to inadequate methodological quality, both of which are more common among small trials. Such bias may distort the results from meta-analyses.[11,12] Bias in randomised controlled trials research has been examined by considering collections of meta-analyses in which component trials are classified according to characteristics such as study quality or publication type. The landmark study by Schulz *et al.*[11] pioneered this approach, for trials with binary outcomes included in meta-analyses from the Cochrane Pregnancy and Childbirth Database. Since then several other studies,[12–19]

have examined the influence of different types of biases on the results of systematic reviews and meta-analyses of randomised controlled trials. Such studies, which are based on meta-regression analyses of several meta-analyses, have been termed 'meta-epidemiological' studies to distinguish them from standard meta-analyses.[20] The applicants include investigators from all identified studies of this type.[12–19,21–23]

These studies have so far identified two dimensions of trial quality that are empirically associated with bias: concealment of the allocation sequence (to prevent selection bias) and blinding (keeping the trial participants, care providers and outcome assessors unaware of which intervention is being administered).[11,12,15] Biased dissemination of the results of trials (publication bias, language bias) may also affect meta-analyses.[24] For none of these factors are the data unequivocal, however:[17] more work is needed to increase confidence in the findings. Further, the relative importance of different sources of bias is unclear at present. A recent comparison of results from these studies indicated that trial quality appears to be more important than the reporting and dissemination of results.[25] However, such comparisons are problematic. First, the definitions of allocation concealment, double blinding etc. varied to some extent across studies. Second, there is a degree of overlap between studies in the meta-analyses included. Third, most studies used logistic regression models in which the evidence for interaction between the effects of trial quality and intervention group is examined, after controlling for the interaction between meta-analysis and treatment group. This assumes that the effect of bias is constant across meta-analyses. If this assumption is false then standard errors of estimated differences will be too small. We have proposed new methods that address these problems, and shown that both within and between meta-analysis heterogeneity may be of importance.[20] These methods require further development, in particular to control for confounding effects of different trial characteristics within a modelling framework that allows appropriately for between meta-analysis variability in the effects of trial characteristics. It would also be desirable to extend this work to different effect measures (for example mean difference, hazard ratio, risk ratio) in addition to odds ratios.

Such influences are likely to vary according to clinical context, but none of the existing studies was large enough to examine this. Specific issues apply in some specialties for which we have limited data, for example surgery and vaccination. Furthermore, much of the literature included in the existing studies is rather old, with a minority of studies from the 1990s. It would be interesting to examine whether improved reporting of trials since the CONSORT statement (Consolidated Standards for the Reporting of Randomised Trials)[26,27] has changed the impact of poor trial quality on meta-analyses, although inevitably the number of post-CONSORT trials included in published meta-analyses is relatively small. In interpreting trends over time it will be important to realise that CONSORT deals mainly with reporting of trials and may have changed reporting behaviour rather than actual conduct. This could lead to the attenuation of apparent associations of trial quality with intervention effect estimates over time.

More information on the epidemiology of bias in randomised clinical trial research is essential to provide a reliable methodological underpinning for unbiased conduct and appropriate interpretation of systematic reviews, meta-analysis and guidelines. Meta-analyses have sometimes been found to produce misleading results.[28,29] At present, unreliable answers are being generated by small randomised studies of doubtful quality and meta-analyses based on such trials give the illusion of reliability.[30] These trials and meta-analyses are nevertheless often the best available evidence, and as such are used to guide decision making in clinical practice. The collaborative research proposed here brings together the leading investigators in the field, and will represent the most comprehensive study ever done in this area. A number of new research questions will be addressed. The results will have considerable impact on clinical trials research, the practice and interpretation of systematic reviews and meta-analysis, and the delivery of effective health care.

Bias in non-randomised studies

Several scenarios remain under which an RCT may be unnecessary, inappropriate, impossible or inadequate.[31] Examples include the assessment of rare side-effects of treatments, some preventive interventions and policy changes. Furthermore, there must be hundreds of examples of interventions for which RCTs would be possible but have not yet been carried out, leaving the medical and policy community to rely on non-randomised evidence only. There are instances where non-randomised studies have either been sufficient to demonstrate effectiveness, or where they appear to have arrived at results similar to those of RCTs. However, where randomisation is possible, most agree that the RCT should be the preferred method of evaluating effectiveness.[32–34] The risks of relying solely on non-randomised evidence include failing to convince some people of the validity of the result, or successfully convincing others of an incorrect result.[33] It would be of value to have clear estimates of the degree to which non-randomised studies may be biased, and whether the bias is consistent across clinical settings, to assist interpretation of non-randomised evidence.

There is inconsistent use of nomenclature when describing non-randomised studies (NRS), and other taxonomies may apply different definitions to the same study designs. To attempt to avoid the problems of inconsistent terminology, six features can be identified that differentiate between these studies. First, some studies make comparisons between groups, whilst some simply describe outcomes in a single group (e.g. case series). Second, comparative designs differ in the way that participants are allocated to groups, varying from the use of randomisation (RCTs), quasi-randomisation, geographical or temporal factors (comparative cohort studies), the decisions of health care professionals (clinical database cohorts), to the identification of comparison groups with specific outcomes (case–control studies). Third, studies differ in the degree to which they are prospective (and therefore planned) or retrospective, for matters such as the recruitment of participants, collection of baseline data, collection of outcome data and generation of hypotheses. Fourth, the method used to investigate comparability of the groups varies: in RCTs no investigation is necessary (although it is often carried out), in controlled before-and-after designs baseline outcome measurements are used, whilst in cohort and case–control studies investigation of confounders is required. Fifth, studies differ in the level at which the intervention is applied: sometimes it is allocated to individuals, other times to groups or clusters. Finally, some studies are classified as experimental whilst others are observational. In experimental studies the study investigator has some degree of control over the allocation of interventions. Most importantly he/she has control over the allocation of participants to intervention groups either using randomisation of participants, or haphazard allocation by alternation, dates of birth, day of the week or case record numbers. In observational studies, on the other hand, the groups that are compared are generated according to variation in the use of interventions that occurs regardless of the study. When allocation is determined largely by health professionals, the treatment decision is based not only on 'hard' data such as age, sex and diagnostic test results, but on 'soft' data including type and severity of symptoms, rate of development of the illness, and severity of any co-morbid conditions, which are rarely made explicit.[35] Allocation in non-randomised studies may also be based on factors such as availability of care or geographical location. In non-randomised studies, therefore, there are likely to be systematic differences in the case-mix of patients in the intervention and comparison groups. Furthermore, some participants may not have been eligible for all the treatments being considered in the study.

The degree to which non-random allocation methods are susceptible to selection bias and therefore may produce biased estimates of the effect of treatment is not clearly understood, although it seems likely that the potential for bias will vary between clinical areas. It is reasonable to expect that the comparability of the groups in terms of prognostic factors, and the extent to which prognosis influences both selection for treatment *and* treatment outcome, will be of particular relevance. For example, in evaluations of childhood vaccines there are few

indicators of prognosis that could be used to influence allocation, so randomisation may not be necessary (although there are dangers that allocation in clusters could be confounded by exposure to infectious disease, which means that in practice randomisation is recommended). By contrast, when patient factors could have a strong influence on allocation or where prognosis is strongly linked to outcomes (such as in cancer treatment), then randomisation is likely to be extremely important.

The use of meta-epidemiology has been informally extended from the comparison of design features of RCTs to comparisons between different study designs. A recent HTA report for which one of the applicants was lead author[36] identified 8 empirical studies[37–44] (7 from the medical field) that compared the results of RCTs with those from non-randomised studies across multiple interventions to estimate the bias removed by randomisation. Between study-design comparisons were noted to be particularly challenging because of the magnitude of potential *meta-confounding* of study design with differences in participants, interventions, outcomes and other features of study design, and because of the likelihood that any bias would act to inflate the variability of estimates as well as acting in a systematic manner.

The conclusions of the eight comparisons of RCTs and NRS are divergent, and all the comparisons were noted to have methodological weaknesses. There are issues as to whether identification of included studies was likely to be biased, the similarity of participants, comparability of interventions and outcome measures between RCTs and NRS, whether study methodology was similar in all respects other than the allocation mechanism, and the definitions used to assess whether results of RCTs and NRS revealed differences or were similar. The only robust conclusion that can be drawn is that in some circumstances the results of randomised and non-randomised studies differ, but it cannot be proved that differences are not due to other confounding factors. The frequency, size and direction of the biases cannot be judged reliably from the information presented in the literature to date. The current study will contribute to clarifying these issues and inform the interpretation of single NRS and systematic reviews and meta-analyses of NRS. The results will also inform the development of guidelines for the reporting of NRS, the Standards for the Reporting of Observational Studies in Epidemiology (STROBE) initiative, which is led by two of the applicants (Matthias Egger, Douglas Altman) and is supported by the Research Methodology Programme.

Aim

To examine the importance of different types of bias and other sources of variability in estimates of the effect of health care interventions, in different settings and areas of medicine.

Objectives

Overall objectives:

1. To implement and enhance recently developed statistical methods to quantify the implications for systematic reviews of the estimated mean and variance of the bias associated with different study characteristics, before and after controlling for confounding between different sources of bias.
2. To examine:
   i. the extent of associations between the different dimensions of methodological quality, publication bias and other reporting biases and the degree of confounding between them
   ii. the relative importance of different biases and study characteristics in terms of distortion of combined effect estimates
   iii. the relative importance of different biases and study characteristics as sources of between-study heterogeneity
   iv. the implications of bias and other sources of between-study heterogeneity for the conduct and analysis of systematic reviews of the effect of health care interventions.

3. To inform the development of improved guidelines for the conduct and reporting of randomised controlled trials and non-randomised studies.

Workstream A: Bias and other sources of variability in the results of randomised controlled trials

1. To extend and analyse a combined database of randomised trials included in meta-analyses that have been assessed according to standard methodological criteria. As part of this process, to refine definitions of blinding and intent to treat analysis, taking into account the feasibility of blinding patients, care givers and outcome assessors, and the explicit reporting, or otherwise, of exclusions from the analysis.

2. For comparisons of randomised controlled trials: to estimate the mean and variance of the bias comparing:
   i. adequately concealed trials with trials where concealment was inadequate and/or remained unclear
   ii. trials using different degrees of blinding with open trials, and whether the importance of blinding varies between trials with 'hard' outcomes (e.g. mortality) and those with more subjective outcomes such as pain relief or patient satisfaction with the treatment regimen.
   iii. trials analysed according to the intent to treat principle with other trials
   iv. mode of generation of the randomisation sequence
   v. trials with varying lengths and completeness of follow up
   vi. trials where the sponsor does and does not have a commercial interest in the results in different fields of medicine, different types of trial and different settings defined by the presence or absence of known prognostic factors. It should be noted that although some of these differences have already been estimated in published studies, there is little or no published information on the between-meta-analysis variation in these differences.

Workstream B: Bias and other sources of variability in the results of non-randomised studies, compared to randomised controlled trials

1. To use a standardised assessment procedure to assemble a new database of sets of randomised trials and non-randomised studies that compared the effect of the same intervention on the same outcome.

2. For comparisons of randomised controlled trials with non-randomised studies estimating the same intervention effect: to estimate the mean and variance of the bias comparing:
   i. randomised and non-randomised studies
   ii. different types of non-randomised studies
   iii. different strategies for dealing with confounding (e.g. logistic regression, propensity score)
   iv. different settings defined by the presence or absence of known prognostic factors (e.g. trials in oncology versus vaccine trials, or paediatric trials, which tend to be particularly small).

3. To inform the development of guidelines for the conduct and reporting of randomised controlled trials and non-randomised studies, and systematic reviews of these studies.

Research methods

The project will involve the construction and analysis of two large databases to be used for empirical research on bias and other sources of variability in estimates of the effect of health care interventions. The first database will contain information on the characteristics and results of randomised controlled trials included in meta-analyses, and will be the responsibility of the research associate to be based in Bristol and directly supervised by Jonathan Sterne (see Workstream A below). The second database will contain information on the characteristics and results of randomised controlled trials and non-randomised studies that evaluated comparable

interventions using comparable outcomes, and will be the responsibility of the research associate to be based in Oxford and directly supervised by Jon Deeks (see Workstream B below). Data analyses, using the methods described in the section on data analysis, will be the responsibility of the project statistician, supervised by Jonathan Sterne, Jon Deeks and Douglas Altman.

Although there appears to be symmetry between Workstreams A and B, in reality there are some important differences. In A we will be looking at effect modifiers (methodological and other) within the class of RCTs to explain variation among RCTs. In B we will examine the magnitude, extent of and reasons for variation between RCTs and NRS, and perhaps also within different types of NRS. Several good studies of effect modifiers in RCTs (the studies in *Table 1*) have already been published, and we propose to build on this work. In contrast, there is (to our knowledge) no published evidence for B. Although there are several studies comparing RCTs and NRS (*Table 2*), none looked in detail at the characteristics of the RCTs and NRS included in those reviews, and there is little information on the magnitude of effect modification comparing RCTs and NRS or between different designs of NRS. Given that Workstream B starts from a much lower knowledge base than Workstream A and also is more complex than Workstream A, the objectives for B are more limited than for A.

Workstream A: Bias and other sources of variability in the results of randomised controlled trials

Identification of relevant studies

Individual patient data (IPD) meta-analyses which involve obtaining individual information or 'raw data' on all patients included in each of the trials included in the review have, in a number of cases, produced definitive answers which might not have been obtained in any other way.[45,46] Similarly, in the realm of empirical studies of bias, a collaborative individual trial data (ITD) meta-analysis will overcome the problems mentioned above by eliminating duplicates, standardising definitions and collaboratively developing an appropriate strategy for the analysis of the common database.

We have identified 11 relevant meta-epidemiological studies of randomised trials: 10 are published[11–19,23] and 1 is ongoing.[21] The lead investigators of all of these studies (characteristics summarised in *Table 1*) have agreed to participate in this project. Meta-analyses included in these studies were mainly identified from Cochrane reviews, from existing databases or from searches of medical journals. Although definitions used in published articles differed it was feasible to establish common definitions for dimensions of trial quality.

Assembly of database for empirical research

Data sets from 7 studies[11,13,15,16,18,19,23] have been combined to form a comprehensive database of systematic reviews and trials. This contains a common set of variables based on agreed definitions. This work has been done in Bristol as part of an MRC-supported PhD project by Mrs Lesley Wood (who has agreed to collaborate on the project); and has been further supported by a grant from the Swiss National Science Foundation.

The unique PMID identifier from PubMed/MEDLINE was assigned to all published meta-analyses and trials from these datasets whenever such an identifier was available. Where there is no PMID the unique identifier from EMBASE was used if available; if not then a unique database ID was assigned. The common database includes 2948 unique trials in 264 unique meta-analyses of which 174 (59%) were extracted from the Cochrane Database of Systematic Reviews. There was surprisingly little duplication: only 363 trials and 23 meta-analyses appeared in two or more of the studies. The original papers for duplicated trials and meta-analyses have been obtained. However, there were differences in the definition of components of trial quality between

meta-epidemiological studies: for example, a paper that described allocation concealment using sealed envelopes without providing further details might or might not be coded as adequately concealed.

We will extend the database to include all the additional studies that have been identified, and aim to continue to maintain and update the database as required for future studies.

Extraction of additional data

We will compile a complete set of the original publications and extract additional information on trials or reviews where current information is inconsistent or incomplete. This process will be informed by recent research and debate on definitions of components of quality such as blinding.[47] For some trial characteristics (e.g. whether published and language of publication) it will be relatively straightforward to add information, while other study characteristics, such as length and completeness of follow up, have not been considered in any of the existing studies and will therefore be more difficult to add. If accumulation of complete data is not cost-effective we will collect information on a subset of trials and/or meta-analyses.

We will identify areas in which the identification and inclusion of additional meta-analyses and trials is required to address our objectives. For example, we may need to identify additional meta-analyses in surgery and of vaccine trials, or to include recently published meta-analyses. We intend to examine at least nine areas in conventional medicine (cardiovascular, oncology, obstetrics and gynaecology, rheumatology, psychiatry, paediatrics, infectious diseases, surgery and vaccine trials). Similarly, we may have to add meta-analyses of trials with certain types of outcome measures or interventions to address some of the objectives. We will adopt a strategic approach and make optimal use of the information available from the existing studies. However, we anticipate that, in addition to collecting additional information on trials and meta-analyses already in the database, we will be able to add approximately 20 meta-analyses and 200 trials to the database.

Investigations

The primary comparison will be between results of RCTs with different characteristics (e.g. components of methodological quality). We will use statistical methods developed by the applicants (see below) to estimate the average effect of the characteristics, the overall variability in these effects and the extent to which such variability is explained by factors such as medical specialty or the presence of known prognostic factors. These analyses will also investigate the extent of confounding between different study characteristics (for example low quality studies may also be less likely to be published). We will quantify the implications of our findings for systematic reviews in which studies of varying quality are located.

Workstream A will be conducted primarily by the research associate based in Bristol, supervised by Jonathan Sterne and Matthias Egger, with additional regular input from Jon Deeks, Douglas Altman and other applicants.

Workstream B: Bias and other sources of variability in the results of non-randomised studies, compared to randomised controlled trials

There are major challenges in investigating differences between the findings of randomised and non-randomised studies. In what follows we outline the issues that we wish to examine, but we realise that it will not be possible to resolve them all. The main impediments are the subjectivity of some of the judgements (e.g. the comparability of interventions in randomised and non-randomised settings) coupled with the quality of the available information in published reports of the primary studies. Thus, the planned analyses will only be carried out if the available data

make them possible. By attempting to address these difficult issues we will help to clarify what is possible and what is not possible in this important area of research.

Identification of relevant studies

Empirical studies comparing randomised and non-randomised studies were identified as part of the published HTA report.[36] Seven studies in medical fields were found, evaluating a total of 82 interventions after accounting for duplication between studies (characteristics summarised in *Table 2*).[37–43] A separate database will be constructed along the same lines as for randomised trials. Here, additional fields will identify the study design (RCT, cohort study, case–control study, etc.) and other characteristics relevant to non-randomised studies. We will search for additional publications that have compared results of RCTs and NRS within multiple meta-analyses, and for other studies that have compared the results of RCTs and NRS.[48] One important and untapped resource for location of such studies is the Evidence-based Practice Center evidence reports (http://www.ahcpr.gov/clinic/epcix.htm). Currently there are 13 EPCs in North America. Their remit is to conduct systematic reviews in a very wide variety of topics. Approximately 90 reviews, which include both randomised and non-randomised studies, have been completed, and these reviews have been able to 'control' (ensure comparability of) the intervention between RCTs and NRS.

Assembly of databases for empirical research

Workstream B will produce a similar common database for the meta-analyses of randomised and non-randomised studies included in the 7 existing empirical reviews of bias associated with non-randomisation. However, as the 7 empirical studies each included and assessed different dimensions of study design, execution and comparability of groups, the original publications for the included randomised and non-randomised studies will each be retrieved and reassessed using a standard protocol. In this work we will draw on the suggestions of Ioannidis and Lau.[49]

A substantial database containing the results of different types of NRS is being constructed during a systematic review of the relationship of diet and physical activity with bladder, prostate, kidney and skin cancer currently underway in Bristol (principal investigators Jonathan Sterne and George Davey Smith). All published information on associations between any aspect of diet or physical activity is being extracted and will be converted to estimates of the association between these exposures and the specified cancer. We have recorded both the type of NRS and key aspects of study quality (for example in case–control studies whether the control group was randomly sampled from a relevant population). Results of the studies are being transformed into common estimates of dose–response effect, according to an algorithm developed by Jonathan Sterne, Matthias Egger and colleagues. This database, which will contain the results of at least 500 NRS, will be used to investigate differences between the results of NRS according to type of study, methodological quality and extent of control for confounding factors.

Extraction of data

An assessment protocol will be developed and its reliability assessed on a sample of studies. Once consensus on the content and wording of the tool is agreed, approximately 300 RCTs and 300 non-randomised studies will be assessed and included in the database of RCTs and NRS. The database of RCTs will be linked with the Workstream A database to check whether any trials have already been assessed.

One area requiring development and detailed analysis will be assessing the similarity of interventions, populations and outcomes, as these may be major sources of confounding which prevent the impact of non-randomisation *per se* being evaluated. For example, a RCT that examined the effect of vitamin E supplementation might be considered comparable to a NRS in which the dominant source of vitamin E was supplements, but not to a NRS in which vitamin E

was mainly of dietary origin. Another issue that is much less simple than for RCTs is to consider dimensions of methodological quality of NRS. The choice of aspects to consider will be informed by the review of Deeks *et al.*[36]

A second question of interest is whether bias in non-randomised allocation is related to the health care context. The most important aspect is probably the degree to which individual prognoses of study participants is predictable. The magnitude of bias that can be introduced in allocation is limited by the accuracy with which prognoses can be made. In studies of primary prevention interventions, such as infant vaccination programs, little prognostic information is known, whereas in studies in the treatment of cancer detailed prognostic information is available. A method of classifying clinical topics according to the amount of prognostic knowledge the researcher can have will be developed, as far as possible blind to knowledge of the comparisons being studied, and used to stratify analyses comparing randomised and non-randomised allocation methods. (The same classification could also be investigated as a possible explanation for inconsistent results of the impact of methodological aspects of RCTs, such as allocation concealment, in Workstream A.) Additional concerns are the difference between alternative experimental and observational methods of constructing non-randomised control groups and the impact of different case-mix adjustment methods in NRS.

Investigations
The primary comparison will be between average results in RCTs with average results in comparable NRS. We will investigate the extent to which candidate predictors of effects of intervention among NRS (e.g. control of confounding, type of study) influence the comparison. We will need to consider the impact of methodological quality of the RCTs in making this comparison. In addition, we will investigate the possibility of greater heterogeneity among NRS after allowance for effect modifiers. A side-product of these analyses will be insight into possible differences between results from different types of NRS (e.g. case–control vs cohort).

With respect to the assessment of the nature of the interventions, two alternative analyses will be undertaken: (a) a restricted analysis where similarity of confounding factors is demonstrated; (b) an unrestricted analysis where all comparisons are included. This approach will allow (a) estimation of biases directly linked to allocation method, and (b) a more pragmatic assessment of the overall differences in findings between randomised and non-randomised evaluations.

The database on studies of diet and cancer will be used to estimate the association between characteristics of NRS, including type of study, methodological quality and extent of control for confounding factors, in the particular context of the association of diet and physical activity with the development of cancer.

We will quantify the implications of our findings for systematic reviews in which estimates of the effect of health interventions are available from both RCTs and NRS.

Workstream B will be conducted primarily by the research associate based in Oxford. Jon Deeks and Douglas Altman will provide day-to-day supervision, with additional regular input from Jonathan Sterne, Matthias Egger and other applicants.

Statistical methods and data analysis
We first describe statistical methods relevant to the analysis of meta-epidemiological studies. We then state the broad aims of the statistical analyses. Predicted outputs from the analyses appear in a separate section.

Standard logistic regression models

Suppose that we have data from M meta-analyses, containing a total of S studies. To estimate the effect of a binary study characteristic C (for example C = 1 in published trials, 0 in unpublished trials) on estimated treatment effects we fit the model:

$$\mathrm{logit}(\pi) = \beta_0 + \beta_1 I_t + \beta_2 I_{tc} + \sum_{i=2}^{M} \gamma_i I_{tm_i} + \sum_{j=2}^{S} \delta_j I_{s_j}$$

where π is the probability that an (adverse) outcome event is observed, I_t, I_{tc}, $\{I_{tm_i}\}$ and $\{I_{s_j}\}$ are all indicator variables denoting, respectively, the effects of treatment, the treatment–characteristic interaction, the treatment–meta-analysis interactions and study number, and $\{\beta\}$, $\{\gamma\}$ and $\{\delta\}$ are the parameters of the logistic regression model.[20] This model allows the probability of the outcome event to vary according to treatment group, trial characteristic and trial, while the interaction terms $\{I_{mi}\}$ mean that the effect of treatment is estimated separately in each meta-analysis. The estimated effect of the characteristic C on average treatment effects is then given by parameter β_2 (treatment–characteristic interaction), which estimates the log of the ratio of treatment odds ratios (ROR) in trials with and without the characteristic. This is assumed to be constant across meta-analyses: violation of this assumption will lead to standard errors that are too small. Only meta-analyses that contain trials with and without the characteristic contribute to this estimate. Corresponding methods using linear regression can be used for meta-analyses in which the outcome is numerical, and mean differences can be converted to corresponding odds ratios to facilitate comparisons across meta-analyses.

Robust standard errors

The assumption that RORs are constant across meta-analyses can be examined via 'robust' standard errors, which use the 'information sandwich'[50,51] to estimate standard errors based on the regression residuals. Robust standard errors may also be estimated after allowing for clustering,[52] providing that the number of clusters is at least 20.[53]

Meta-analytic approaches

If the effect of a trial characteristic varies between meta-analyses then analyses based on the logistic regression approach will underestimate the uncertainty in estimated RORs. An obvious alternative is to estimate the effect of the characteristic using a separate logistic regression in each meta-analysis (i.e. fixed-effect within meta-analyses). Estimated RORs in each meta-analysis can then be combined using meta-analytic methods, using inverse-variance weighting and either fixed-effect or random-effects between meta-analyses. Random-effects analyses use the moment-based variance estimator $(\hat{\tau}^2)$ proposed by DerSimonian and Laird.[54] The fixed-effect assumption may also be violated within meta-analyses. This can be addressed by using random-effects meta-regression to allow for between-trial (within meta-analysis) heterogeneity.[55,56] Meta-regression examines associations between the estimated treatment effect (log OR) in each trial and one or more trial characteristics, allowing appropriately for the precision of the treatment effect via the standard error of the log OR in each trial. For a single meta-analysis, meta-regression estimates the same quantity (the ratio of ORs comparing trials with and without the characteristic) as is estimated using the logistic regression approach. In the absence of within meta-analysis heterogeneity the ratios of ORs estimated using meta-regression and logistic regression will be similar. Estimates of the between meta-analysis variance in the effect of trial characteristics are likely to be of key importance in synthesis of evidence from studies of different types (see 'Evidence synthesis' below).

Random effects logistic regression and multilevel modelling

We have experience with all the methods described above[20] but have not yet used two promising alternatives: first, the logistic regression approach could be extended to allow for random effects

between or within meta-analyses and, second, a multilevel approach could be developed, based on summary treatment effect estimates from each trial.[57] As well as providing estimates of both the effect of trial characteristics and the between meta-analysis variance in these effects, these approaches could have particular advantages in analyses that try to control for confounding effects of other trial characteristics, since the estimated effect of a characteristic in a particular meta-analysis would be shrunk towards the overall mean by an amount depending on the amount of information in the data for that meta-analysis. In analyses using the meta-analytic approach we have seen that the effect of trial characteristics is estimated very imprecisely, which will make it difficult to control for confounding effects. The logistic regression approach discussed above could obviously be used to control for confounding (for example between different components of methodological quality) by including the effects of more than one trial characteristic in the model. However, this assumes that the effect of each characteristic is constant across meta-analyses. An alternative is to estimate effects controlling for confounding factors separately in each meta-analysis, using random-effects meta-regression.

Evidence synthesis using recently developed methodology

Jonathan Sterne (lead applicant) and John Carlin (Professor, University of Melbourne Departments of Paediatrics and Public Health) have recently been developing methods to use the results of meta-epidemiological studies to provide estimates of the effect of interventions based on evidence from studies of differing methodological quality, or different types of study. We now briefly describe this: a paper is currently being prepared for publication.

We wish to combine the results of two types of studies to provide an overall estimate of the effect of a particular intervention. We will denote these two types of studies by H (high quality) and L (low quality). Two examples are that H and L could be RCTs in which randomisation was and was not adequately concealed, or that H could be RCTs while L could be cohort studies.

We will assume that meta-analyses of the effect of the intervention are conducted separately in studies of the two types, leading to estimates $\hat{\beta}_H$ and $\hat{\beta}_L$ of the intervention effect. For meta-analyses based on binary outcomes, $\hat{\beta}_H$ and $\hat{\beta}_L$ will generally correspond to the log intervention odds ratios in type H and type L studies respectively. The variances of $\hat{\beta}_H$ and $\hat{\beta}_L$ will be denoted by σ_H^2 and $\sigma_L^2 >$ respectively. Based on the usual assumption that intervention effect estimates from meta-analysis are asymptotically normally distributed, we will assume that $\hat{\beta}_H \sim N(\mu, \sigma_H^2)$ ($\hat{\beta}_H$ is normally distributed with mean μ and variance σ_H^2). We will interpret μ as the (true) intervention effect in type H studies, where this is assumed to be constant across all studies (the usual 'fixed effect' assumption).

We will assume that in type L studies, the intervention effect is estimated with bias δ (in studies with binary outcomes, δ is the log of the ratio of intervention odds ratios comparing type L and type H studies). Therefore, $\hat{\beta}_L \sim N(\mu + \delta, \sigma_L^2)$. If δ were known, then we could obtain a corrected estimate of μ, based on the type L studies, as $\hat{\mu}_L = \hat{\beta}_L - \delta$. However, if δ is unknown, as will invariably be the case, the results of the type L studies cannot distinguish μ and δ without further assumptions.

A solution to this problem is to incorporate prior information about the bias δ, using a Bayesian framework. As described earlier, the effect of components of trial quality on intervention effect estimates, and the between meta-analysis variability in such effects, may be estimated in meta-epidemiological studies, using data from collections of meta-analyses. For example, in a recent

re-analysis of the data of Schulz *et al.*[11] the ratio of odds ratios (ROR) comparing studies that were not and were adequately concealed was 0.67 (95% CI 0.57 to 0.78), while the between-meta-analysis variance in the log ROR was 0.065.[20] We propose to formalise evidence about this variability in a prior distribution for δ, and a convenient assumption is a normal distribution: $\delta \sim N(\delta_0, \tau_0^2)$, where δ_0 represents our best *a priori* estimate of the average bias in studies of type L, while τ_0^2 describes uncertainty around this estimate. The results just quoted could motivate the particular specifications $\delta_0 = \log_e 0.67 = -0.40$ and $\tau_0^2 = 0.1$, say.

Based on these assumptions, it is straightforward to obtain the posterior distribution of the true intervention effect μ, given the information from both type H and type L studies. In particular, integrating over the prior distribution of δ, we obtain the marginal distribution of $\hat{\beta}_L \sim N(\mu + \delta_0, \sigma_L^2 + \tau_0^2)$.

It follows by standard normal distribution calculations,[58] assuming a diffuse ('noninformative') prior distribution for μ, that the posterior distribution for μ given the combined data $D = (\hat{\beta}_H, \hat{\beta}_L)$ is normal with mean and variance

$$E(\mu \mid D) = \frac{\dfrac{\hat{\beta}_H}{\sigma_H^2} + \dfrac{\hat{\beta}_L - \delta_0}{\sigma_L^2 + \tau_0^2}}{\dfrac{1}{\sigma_H^2} + \dfrac{1}{\sigma_L^2 + \tau_0^2}} \tag{1}$$

$$Var(\mu \mid D) = \left[\frac{1}{\sigma_H^2} + \frac{1}{\sigma_L^2 + \tau_0^2} \right]^{-1} \tag{2}$$

Equations 1 and 2 show that the posterior mean of μ is a weighted average of the estimates of μ from the two types of studies ($\hat{\beta}_H$ in the type H studies and $\hat{\beta}_L - \delta_0$ in the type L studies), with weights $[\sigma_H^2]^{-1}$ and $[\sigma_L^2 + \tau_0^2]^{-1}$ in the type H and type L studies respectively. It follows that τ_0^2, the prior variability in the amount of bias, is key in determining the contribution of the type L studies to the posterior mean of μ, since the larger is τ_0^2, the smaller is the weight given to the type L studies. When τ_0^2 is zero (the amount of bias is known with certainty), equations 1 and 2 correspond to a standard inverse-variance weighted meta-analysis of the estimates of μ from the two types of studies. As τ_0^2 tends to infinity, the contribution of the type L studies tends to zero, and equations 1 and 2 simplify to give $E(\mu) = \hat{\beta}_H$, $Var(\mu) = \sigma_H^2$, so that the posterior information about μ corresponds to the information provided by the type H studies. Note also that the information from the type L studies is both corrected for bias and downweighted according to our uncertainty about the magnitude of the bias.

These statistical methods will be used to analyse data from workstreams A and B, during the second 12 months of the project. We will estimate both the mean differences in estimated effect of interventions, and the variation in these differences, according to characteristics of studies and types of studies, before and after controlling for the confounding effects of other study characteristics. This will lead to formal proposals, based on the methods for evidence synthesis described above on how to combine estimates of the effect of interventions from RCTs differing in quality and from different types of study.

Schedule of work

Task	Month											
	1–2	3–4	5–6	7–8	9–10	11–12	13–14	15–16	17–18	19–20	21–22	23–24
Workstream A												
Obtain data from additional meta-epidemiological studies	•											
Combine new studies with existing database		•	•									
Retrieve papers and extract new information on trials		•	•	•	•	•	•	•	•			
Workstream B												
Design, piloting and revision of data extraction forms	•	•										
Retrieve original publications	•	•	•	•	•	•						
Data extraction			•	•	•	•	•	•	•			
Analysis and reporting												
Statistical analysis of data from both workstreams							•	•	•	•	•	•
Preparation of reports to funder and papers for publication										•	•	•

Outputs

Workstream A

- Updated estimates of the effect of components of trial quality and reporting characteristics on estimated treatment effects in randomised controlled trials, both overall and for the first time for different clinical settings, types of intervention, outcomes and trial designs.
- Refined definitions of established markers of trial quality and identification of additional markers, including aspects of reporting (reporting of sample size calculations, eligibility criteria, primary and secondary outcomes), the degree of baseline imbalances, the number of centres involved, and the source of funding.
- An improved understanding of the extent and effect of confounding between dimensions of methodological quality and other trial characteristics, and hence their relative importance in the epidemiology of bias in clinical trial research and the extent to which confounding by different trial characteristics affects estimates of the effect of methodological quality and reporting biases.
- New estimates of the effect and impact of components of trial quality in different medical specialties, comparing conventional with complementary medicine, and comparing settings defined by the presence or absence of known prognostic factors.
- New estimates of the effect and impact of components of trial quality in different types of controlled clinical trials, and comparing trials with 'hard' and 'soft' end points.

Workstream B

- Estimates of the average difference between estimates of health interventions derived from RCTs and NRS.
- Estimates of the importance of prognostic information in the difference between randomised and non-randomised studies.
- Estimates of the difference between alternative experimental and observational methods of constructing non-randomised control groups.
- Estimates of the impact of case-mix adjustment methods in non-randomised studies.

General

- All analyses will lead to estimates of the between meta-analysis variation in the effect of study characteristics on intervention effect estimates. These will be used in analyses showing the effect of combining different types of studies in meta-analyses, based on the methods for evidence synthesis described earlier and incorporating prior estimates of both the expected bias and its variance.
- Based on these results, and on refined methods of characterising trials developed as part of Workstream A, we will produce updated recommendations for the conduct, reporting and appraisal of randomised controlled trials, including revised editions of the CONSORT statement and the continuing development of the STROBE statement. All applicants will contribute to planning, conducting, interpreting and reporting these analyses. We will also aim to contribute to the continuing development of guidelines for the conduct and reporting of systematic reviews and meta-analyses, including the QUORUM statement, particularly to the debate over whether and how to include studies of differing methodological quality and/or different study designs.
- Ultimately, the project will contribute to improved interpretation of existing evidence about the effect of medical interventions, the improved conduct and reporting of future randomised controlled trials and other evaluations of the effect of medical interventions, and hence to better medical care.

People

The team of applicants comprises the leading investigators in the field. All have extensive experience of conducting randomised trials and meta-analyses. Together, the team has links with clinical researchers in many branches of medicine, and covers a breadth of expertise, including clinical medicine and epidemiology (Egger, Ioannidis, Gluud, Pildal), methodological and reporting issues in randomised trials research (Altman, Schulz, Moher, Egger, Ioannidis), Cochrane methodology and Cochrane databases (Gluud, Ioannidis, Jüni, Pildal) and the statistical issues involved in the analysis of meta-epidemiological studies (Sterne, Altman, Ioannidis, Deeks, Schulz). Several applicants (Moher, Altman, Schulz, Egger, Ioannidis) are members of the CONSORT Group,[26,27] the QUOROM Group (Moher, Altman) and the STROBE Group (Egger, Sterne, Altman) and most are affiliated to the Cochrane Collaboration. Three of the applicants (Egger, Moher, Sterne) are co-convenors of the Cochrane Reporting Bias Methods Group and two (Altman, Deeks) are co-convenors of the Cochrane Statistical Methods Group.

Dissemination of research results

We will disseminate our findings in journal articles and at national and international meetings. The final project report will be made available to organisations in the UK and worldwide that fund trials and systematic reviews. In particular, we will meet with representatives of the NHS Health Technology Assessment (HTA) programme and the National Institute for Clinical Excellence (NICE) to discuss the implications of our findings for the research that they commission. We will prepare educational articles and materials, and will make particular efforts to disseminate these in specialist areas for which the research identifies problems with the conduct and reporting of trials. A number of the applicants and collaborators run regular workshops at meetings of the Cochrane Collaboration, and these, together with courses on randomised trials and systematic reviews run in Bristol and Oxford, will incorporate the results of the research as it becomes available. We will also discuss, with other colleagues involved in the Cochrane Collaboration, how the Cochrane Reviewers' Handbook might be updated in the light of our results. Those applicants involved with the development and updating of the CONSORT statement on the reporting of randomised controlled trials will ensure that this is updated to take account of new empirical evidence on aspects of trial quality which emerges from the study. Perhaps the most powerful influence on the conduct and reporting of trials and systematic reviews is medical journals' publication policies. Here, CONSORT has direct and indirect influence: direct through the membership of several editors of the most influential journals and

indirect through endorsement from International Committee of Medical Journal Editors and Council of Science Editors. Similarly, the findings of this research will feed into updates of and extensions to the QUOROM and STROBE Statements.

References

1. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. In Bailar JC, Mosteller F, editors. *Medical uses of statistics*. 2nd edn. Boston, MA: NEJM Books; 1992. pp. 357–73.

2. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;**272**(2):122–4.

3. Thornley B, Adams C. *Content and quality of* 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;**317**:1181–4.

4. McDonald S, Westby M, Clarke M, Lefebvre C and the Cochrane Centres' Working Group. *Number and size of randomised trials reported in general health care journals from* 1948 to 1997. *Int J Epidemiol* 2002;**31**:125–7.

5. Chan AW, Altman DG. A cross-sectional study of randomised controlled trials published in 2000. *submitted* 2003.

6. Egger M, Smith DG, O'Rourke K. Rationale, potentials and promise of systematic reviews. In Egger M, Smith DG, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books; 2001. pp. 23–42.

7. Eccles M, Freemantle N, Mason J. Using systematic reviews in clinical guideline development. In Egger M, Smith GD, Altman DG, editors. Systematic Reviews in Health Care: Meta-Analysis in Context. London: BMJ Books; 2001. 400–409.

8. Gray JAM. Using Systematic Reviews for Evidence Based Policy Making. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books; 2001. pp. 410–18.

9. Sterne JAC, Gavaghan DJ, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53**:1119–29.

10. Mallet S, Clarke M. The typical Cochrane review. How many trials? How many participants? *Int J Technol Assess Health Care* 2002;**18**:820–31.

11. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**(5):408–12.

12. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.

13. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;**356**:1228–31.

14. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, *et al.* What contributions do languages other than English make on the results of meta-analyses. *J Clin Epidemiol* 2000;**53**(9):964–72.

15. Kjaergard LL, Villumsen J, Gluud C. Reported methodological quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;**135**:982–9.

16. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;**31**(1):115–23.

17. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, *et al.* Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;**287**(22):2973–82.

18. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;**7**(1).

19. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, *et al.* Should meta-analysts search EMBASE in addition to MEDLINE? *J Clin Epidemiol* 2003;**56**(10):943–55.

20. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21**:1513–24.

21. Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Bradburn MJ, Altman DG, *et al.* How often do positive conclusions drawn from meta-analyses remain substantiated if only data from randomised trials with adequate allocation concealment are considered? Presentation at 12th Cochrane Colloquium, Ottawa, 2004.

22. Contopoulos-Ioannidis D, Gilbody S, Trikalinos TA, Churcill R, Wahlbeck K, Ioannidis JP. Comparison of large vs. smaller randomized trials for mental health related interventions. *Am J Psychiatry* 2005;**162**:578–84.

23. Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care* 2003;**19**(4):591–603.

24. Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *BMJ* 1998;**316**:61–6.

25. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;**7**(1).

26. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**(8):663–94.

27. Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet* 2001;**357**:1191–4.

28. LeLorier J, Grégoire 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.

29. Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In Egger M, Smith DG, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books; 2001. pp. 43–68.

30. Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998;**317**:1170–1.

31. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**(7040):1215–18.

32. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;**52**(6):487–97.

33. Green SB, Byar DP. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med* 1984;**3**:361–70.

34. Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med* 1997;**90**(7):379–86.

35. Moses LE. Measuring effects without randomized trials? Options, problems, challenges. *Med Care* 1995;**33**(4 Suppl.):AS8–14.

36. 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).

37. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;**2**(13).

38. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;**4**(34).

39. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med* 1982;**72**:233–40.

40. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. The Cochrane Library, Issue 4. Oxford: Update Software; 2002.

41. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**(25):1878–86.

42. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**(25):1887–92.

43. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, *et al.* Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;**286**(7):821–30.

44. Wilson DB, Lipsey MW. The role of method in treatment effectiveness research: evidence from meta-analysis. *Psychol Methods* 2001;**6**:413–29.

45. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;**341**:418–22.

46. Horton R. The information wars [see comments]. *Lancet* 1999;**353**(9148):164–5.

47. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, *et al.* Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;**285**(15):2000–3.

48. Bhandari M, Tornetta P, III, Ellis T, Audige L, Sprague S, Kuo JC, *et al.* Hierarchy of evidence: differences in results between non-randomized studies and randomized trials in patients with femoral neck fractures. *Arch Orthop Trauma Surg* 2004;**124**(1):10–16.

49. Ioannidis JP, Lau J. Heterogeneity of the baseline risk within patient populations of clinical trials: a proposed evaluation algorithm. *Am J Epidemiol* 1998;**148**(11):1117–26.

50. Huber PJ. The behaviour of maximum likelihood estimates under non-standard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 1st edn. Berkeley, CA: University of California; 1967. pp. 221–33.

51. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980;**48**:817–30.

52. Rogers WH. sg17: regression standard errors in clustered samples. *Stata Technical Bulletin* 1993;**3**:88–94.

53. Donner A. Some aspects of the design and analysis of cluster randomization trials. *Appl Statist* 1998;**47**(1):95–113.

54. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.

55. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med* 1995;**14**(4):395–411.

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

57. 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**(24):3417–32.

58. Gelman A, Carlin JB, Stern HB, Rubin DB. *Bayesian data analysis*. 2nd edn. New York: Chapman and Hall; 2003.

TABLE 1 Overview of empirical studies of bias in randomised controlled trials

Study	Main comparison	Clinical topic	No. of meta-analyses	Trials appearing in >1 component meta-analysis excluded?	No. of trials	Year of trial publication	Source of meta-analyses
Schulz et al.[11]	Study quality (allocation concealment, randomisation sequence generation method, blinding and exclusion of patients)	Pregnancy & childbirth	33	Yes, if several overlapping trials. For minor overlap a single trial was removed	250	1955–1992	Reviews from Cochrane Pregnancy & Childbirth Group with ≥5 trials containing ≥25 events in the control group, and at ≥1 trial with and one without adequate allocation concealment
Moher et al.[12]	Study quality (allocation concealment method, randomisation sequence generation method, blinding and exclusion of patients)	Circulatory & digestive, mental health, pregnancy & childbirth	11	Yes	127	1960–1995	Random selection of meta-analyses from large database
Kjaergard et al.[15]	Large and small trials according to quality (sequence generation, concealment, blinding and exclusion of patients)	Various, including cardiac, childbirth, schizophrenia, smoking cessation	14	Yes, overlapping meta-analyses were excluded	190	1960–1998	Meta-analyses in the Cochrane Library, MEDLINE or PubMed with at least one trial with more than 1000 patients
McAuley et al.,[13] Moher et al.[14]	Grey literature, and trials published in languages other than English	Various, including GI, cardiac, infection, reproduction, circulatory	41 (grey literature), 18 (language)	Yes	467 (grey literature), 178 (language)	Min. range 1984–1990	Random selection of meta-analyses from large database
Jüni et al.,[16] Egger et al.[18]	Grey literature, non-English languages, indexing in MEDLINE, study quality (concealment of allocation and blinding)	Any	39 to 60	Yes, trials that appeared in >1 meta-analysis were counted only once	From 304 (concealment) to 783 (grey literature)	1955–1998	Meta-analyses from 8 specified journals, HTA reports, DARE and Cochrane Database of Systematic Reviews that had performed comprehensive literature searches
Balk et al.[17]	Study quality (including 28 distinct quality measures)	Cardiovascular disease, infectious diseases, pediatrics, surgery	26	No overlap	276	1955–2000	MEDLINE and Cochrane Database of Systematic Reviews; for cardiovascular diseases, database from Lau et al. NEJM 1992
Contopoulos-Ioannidis et al.[22]	Study quality (allocation concealment, blinding, sequence generation), also compared large and small trials	Mental health	16	For overlapping reviews, retained the latest with most complete information	133	1960–2002	Cochrane Database of Systematic Reviews, DARE, Mental Health Library (415 reviews screened)
Royle[23]	Indexing in major bibliographic databases and (a) concealment method and (b) number of patients enrolled	Any	29	Meta-analyses not examined for duplicate trials	541	Not stated	Cochrane reviews new to 2001 Issue 1 that included at least 1 unindexed trial
Pildal (unpublished)	The % conclusions that remain supported if only adequately concealed trials are considered. (Data on blinding & sequence generation are available)	Any	70 (34 allowed subgroup analysis)	3 trials re-occurred in two meta-analyses that comprised 6 and 17 each but addressed different outcomes	499	Pending, will be available	Recent reviews stating a conclusive preference supported by a statistically significant binary outcome. 32 were from PubMed and 38 from Cochrane Library

TABLE 2 Overview of empirical studies of bias in non-randomised studies

	Sacks	Kunz	Britton	MacLehose	Benson	Concato	Ioannidis
Comparisons	6	23 (11)	18	14 (38 outcomes)	19	5	45
Non-randomised study designs included	Historically controlled trials (HCTs)	Quasi-experiments, historically controlled trials (HCTs), patient preference trials	Quasi-experiments, natural experiments and prospective observational studies	Quasi-experimental and observational studies	Observational studies	Case–control and cohort studies	Quasi-experiments, cohort and case–control studies
Number of RCTs/NRS	50/56	263/246	46/41	31/68	83/53	55/44	240/168
Method of identification of comparisons	*Primary studies:* Personal systematic collection of RCTs and HCTs in specific fields of interest	*Secondary studies:* Electronic and manual search for studies comparing randomised and non-randomised studies	*Primary and secondary studies:* Electronic searches for studies comparing randomised and non-randomised groups	*Primary and secondary studies:* Electronic and manual searches for studies comparing randomised and non-randomised groups	*Primary studies:* Electronic search of MEDLINE and CCTR for observational studies and matching RCTs	*Secondary studies:* Meta-analyses published in 5 leading journals that included non-randomised studies	*Secondary studies:* Meta-analyses located by electronic searches (MEDLINE and Cochrane library) and manual searches
Similarity of patients, interventions and outcomes	Same intervention for the same medical condition	2 controlled for clinical differences in participants and interventions, 6 did so partly, and 7 did not at all.	Same intervention, similar setting, similar control therapy, comparable outcomes measures	Same intervention. Similarity of eligibility, time period, co-morbidity, disease severity and other prognostic factors assessed	Restricted to interventions allocated by physicians to induce comparability	Not assessed	Not assessed
Similarity of other study methods	Not assessed	5 papers judged partly comparable, 10 not comparable on double blinding, complete follow-up and other methodological issues	Not assessed	Blinding of outcome assessed	Not assessed	Analysis according to study quality mentioned in discussion, but no methods or results presented	Not assessed
Method of summarising study findings	Vote counting of results classified as positive (either statistically significant or had positive conclusions if no statistical analysis)	No consistent summary; results of randomised and non-randomised groups described using a variety of measures	Results of randomised and non-randomised groups described using risk differences, risk ratios or odds ratios	Risk ratios and risk differences calculated separately for MAs for randomised and non-randomised studies	Calculation of overall meta-analytical results and confidence intervals (odds ratios for binary data, differences in means for continuous)	Calculation of overall meta-analytical results and confidence intervals (risk ratios and odds ratios)	Calculation of fixed and random meta-analytical estimates expressed as odds ratios and log odds ratios
Method for comparing results between groups	Comparison of the percentage with positive results	No overall analysis presented	Statistical significance of the difference in effect sizes	Distribution of relative and absolute differences in results reported	Assessment of whether observational point estimate fell within 95% CI for RCTs	No overall analysis presented	Calculation of Z-scores for difference between treatment effects
Criteria for comparing variability of results	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Dispersion of points calculated without considering differences in sample size	Significance ($p < 0.1$) of tests of between-study heterogeneity

Health Technology Assessment programme

Director,
Professor Tom Walley, CBE,
Director, NIHR HTA programme,
Professor of Clinical Pharmacology,
Department of Pharmacology and Therapeutics,
University of Liverpool

Deputy Director,
Professor Hywel Williams,
Professor of Dermato-Epidemiology,
Centre of Evidence-Based Dermatology,
University of Nottingham

Prioritisation Group

Members

Chair,
Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, Department of
Pharmacology and Therapeutics,
University of Liverpool

Professor Imti Choonara,
Professor in Child Health,
Academic Division of Child
Health, University of Nottingham
Chair – Pharmaceuticals Panel

Dr Bob Coates,
Consultant Advisor – Disease
Prevention Panel

Dr Andrew Cook,
Consultant Advisor – Intervention
Procedures Panel

Dr Peter Davidson,
Director of NETSCC, Health
Technology Assessment

Dr Nick Hicks,
Consultant Adviser – Diagnostic
Technologies and Screening Panel,
Consultant Advisor–Psychological
and Community Therapies Panel

Ms Susan Hird,
Consultant Advisor, External
Devices and Physical Therapies
Panel

Professor Sallie Lamb,
Director, Warwick Clinical Trials
Unit, Warwick Medical School,
University of Warwick
Chair – HTA Clinical Evaluation
and Trials Board

Professor Jonathan Michaels,
Professor of Vascular Surgery,
Sheffield Vascular Institute,
University of Sheffield
Chair – Interventional Procedures
Panel

Professor Ruairidh Milne,
Director – External Relations

Dr John Pounsford,
Consultant Physician, Directorate
of Medical Services, North Bristol
NHS Trust
Chair – External Devices and
Physical Therapies Panel

Dr Vaughan Thomas,
Consultant Advisor –
Pharmaceuticals Panel, Clinical
Lead – Clinical Evaluation Trials
Prioritisation Group

Professor Margaret Thorogood,
Professor of Epidemiology, Health
Sciences Research Institute,
University of Warwick
Chair – Disease Prevention Panel

Professor Lindsay Turnbull,
Professor of Radiology, Centre for
the MR Investigations, University
of Hull
Chair – Diagnostic Technologies
and Screening Panel

Professor Scott Weich,
Professor of Psychiatry, Health
Sciences Research Institute,
University of Warwick
Chair – Psychological and
Community Therapies Panel

Professor Hywel Williams,
Director of Nottingham Clinical
Trials Unit, Centre of Evidence-
Based Dermatology, University of
Nottingham
Chair – HTA Commissioning
Board
Deputy HTA Programme Director

HTA Commissioning Board

Chair,
Professor Hywel Williams,
Professor of Dermato-Epidemiology,
Centre of Evidence-Based Dermatology,
University of Nottingham

Deputy Chair,
Professor Jon Deeks,
Professor of Bio-Statistics,
Department of Public Health and
Epidemiology,
University of Birmingham

Programme Director,
Professor Tom Walley, CBE,
Professor of Clinical Pharmacology,
Department of Pharmacology and Therapeutics,
University of Liverpool

Members

Professor Zarko Alfirevic,
Head of Department for Women's
and Children's Health, Institute of
Translational Medicine, University
of Liverpool

Professor Judith Bliss,
Director of ICR-Clinical Trials
and Statistics Unit, The Institute of
Cancer Research

Professor David Fitzmaurice,
Professor of Primary Care
Research, Department of Primary
Care Clinical Sciences, University
of Birmingham

Professor John W Gregory,
Professor in Paediatric
Endocrinology, Department of
Child Health, Wales School of
Medicine, Cardiff University

Professor Steve Halligan,
Professor of Gastrointestinal
Radiology, Department of
Specialist Radiology, University
College Hospital, London

Professor Angela Harden,
Professor of Community and
Family Health, Institute for
Health and Human Development,
University of East London

Dr Joanne Lord,
Reader, Health Economics
Research Group, Brunel University

Professor Stephen Morris,
Professor of Health Economics,
University College London,
Research Department of
Epidemiology and Public Health,
University College London

Professor Dion Morton,
Professor of Surgery, Academic
Department of Surgery, University
of Birmingham

Professor Gail Mountain,
Professor of Health Services
Research, Rehabilitation and
Assistive Technologies Group,
University of Sheffield

Professor Irwin Nazareth,
Professor of Primary Care and
Head of Department, Department
of Primary Care and Population
Sciences, University College
London

Professor E Andrea Nelson,
Professor of Wound Healing and
Director of Research, School of
Healthcare, University of Leeds

Professor John David Norrie,
Director, Centre for Healthcare
Randomised Trials, Health
Services Research Unit, University
of Aberdeen

Professor Barney Reeves,
Professorial Research Fellow
in Health Services Research,
Department of Clinical Science,
University of Bristol

Professor Peter Tyrer,
Professor of Community
Psychiatry, Centre for Mental
Health, Imperial College London

Professor Martin Underwood,
Professor of Primary Care
Research, Warwick Medical
School, University of Warwick

HTA Commissioning Board *(continued)*

Professor Caroline Watkins,
Professor of Stroke and Older
People's Care, Chair of UK
Forum for Stroke Training, Stroke
Practice Research Unit, University
of Central Lancashire

Dr Duncan Young,
Senior Clinical Lecturer and
Consultant, Nuffield Department
of Anaesthetics, University of
Oxford

Observers

Dr Tom Foulks,
Medical Research Council

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

HTA Clinical Evaluation and Trials Board

Chair,
Professor Sallie Lamb,
Director,
Warwick Clinical Trials Unit,
Warwick Medical School,
University of Warwick and Professor of
Rehabilitation,
Nuffield Department of Orthopaedic,
Rheumatology and Musculoskeletal Sciences,
University of Oxford

Deputy Chair,
Professor Jenny Hewison,
Professor of the Psychology of Health Care,
Leeds Institute of Health Sciences,
University of Leeds

Programme Director,
Professor Tom Walley, CBE,
Director, NIHR HTA programme,
Professor of Clinical Pharmacology,
University of Liverpool

Members

Professor Keith Abrams,
Professor of Medical Statistics,
Department of Health Sciences,
University of Leicester

Professor Martin Bland,
Professor of Health Statistics,
Department of Health Sciences,
University of York

Professor Jane Blazeby,
Professor of Surgery and
Consultant Upper GI Surgeon,
Department of Social Medicine,
University of Bristol

Professor Julia M Brown,
Director, Clinical Trials Research
Unit, University of Leeds

Professor Alistair Burns,
Professor of Old Age Psychiatry,
Psychiatry Research Group, School
of Community-Based Medicine,
The University of Manchester &
National Clinical Director for
Dementia, Department of Health

Dr Jennifer Burr,
Director, Centre for Healthcare
Randomised trials (CHART),
University of Aberdeen

Professor Linda Davies,
Professor of Health Economics,
Health Sciences Research Group,
University of Manchester

Professor Simon Gilbody,
Prof of Psych Medicine and Health
Services Research, Department of
Health Sciences, University of York

Professor Steven Goodacre,
Professor and Consultant in
Emergency Medicine, School of
Health and Related Research,
University of Sheffield

Professor Dyfrig Hughes,
Professor of Pharmacoeconomics,
Centre for Economics and Policy
in Health, Institute of Medical
and Social Care Research, Bangor
University

Professor Paul Jones,
Professor of Respiratory Medicine,
Department of Cardiac and
Vascular Science, St George's
Hospital Medical School,
University of London

Professor Khalid Khan,
Professor of Women's Health and
Clinical Epidemiology, Barts and
the London School of Medicine,
Queen Mary, University of London

Professor Richard J McManus,
Professor of Primary Care
Cardiovascular Research, Primary
Care Clinical Sciences Building,
University of Birmingham

Professor Helen Rodgers,
Professor of Stroke Care, Institute
for Ageing and Health, Newcastle
University

Professor Ken Stein,
Professor of Public Health,
Peninsula Technology Assessment
Group, Peninsula College
of Medicine and Dentistry,
Universities of Exeter and
Plymouth

Professor Jonathan Sterne,
Professor of Medical Statistics
and Epidemiology, Department
of Social Medicine, University of
Bristol

Mr Andy Vail,
Senior Lecturer, Health Sciences
Research Group, University of
Manchester

Professor Clare Wilkinson,
Professor of General Practice and
Director of Research North Wales
Clinical School, Department of
Primary Care and Public Health,
Cardiff University

Dr Ian B Wilkinson,
Senior Lecturer and Honorary
Consultant, Clinical Pharmacology
Unit, Department of Medicine,
University of Cambridge

Observers

Ms Kate Law,
Director of Clinical Trials,
Cancer Research UK

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies and Screening Panel

Members

Chair,
Professor Lindsay Wilson Turnbull,
Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary

Professor Judith E Adams, Consultant Radiologist, Manchester Royal Infirmary, Central Manchester & Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester

Mr Angus S Arunkalaivanan, Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham

Dr Diana Baralle, Consultant and Senior Lecturer in Clinical Genetics, University of Southampton

Dr Stephanie Dancer, Consultant Microbiologist, Hairmyres Hospital, East Kilbride

Dr Diane Eccles, Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital

Dr Trevor Friedman, Consultant Liason Psychiatrist, Brandon Unit, Leicester General Hospital

Dr Ron Gray, Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford

Professor Paul D Griffiths, Professor of Radiology, Academic Unit of Radiology, University of Sheffield

Mr Martin Hooper, Public contributor

Professor Anthony Robert Kendrick, Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton

Dr Nicola Lennard, Senior Medical Officer, MHRA

Dr Anne Mackie, Director of Programmes, UK National Screening Committee, London

Mr David Mathew, Public contributor

Dr Michael Millar, Consultant Senior Lecturer in Microbiology, Department of Pathology & Microbiology, Barts and The London NHS Trust, Royal London Hospital

Mrs Una Rennard, Public contributor

Dr Stuart Smellie, Consultant in Clinical Pathology, Bishop Auckland General Hospital

Ms Jane Smith, Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds

Dr Allison Streetly, Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine

Dr Matthew Thompson, Senior Clinical Scientist and GP, Department of Primary Health Care, University of Oxford

Dr Alan J Williams, Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital

Observers

Dr Tim Elliott, Team Leader, Cancer Screening, Department of Health

Dr Joanna Jenkinson, Board Secretary, Neurosciences and Mental Health Board (NMHB), Medical Research Council

Professor Julietta Patnick, Director, NHS Cancer Screening Programme, Sheffield

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health

Disease Prevention Panel

Members

Chair,
Professor Margaret Thorogood,
Professor of Epidemiology, University of Warwick Medical School, Coventry

Dr Robert Cook, Clinical Programmes Director, Bazian Ltd, London

Dr Colin Greaves, Senior Research Fellow, Peninsula Medical School (Primary Care)

Mr Michael Head, Public contributor

Professor Cathy Jackson, Professor of Primary Care Medicine, Bute Medical School, University of St Andrews

Dr Russell Jago, Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol

Dr Julie Mytton, Consultant in Child Public Health, NHS Bristol

Professor Irwin Nazareth, Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London

Dr Richard Richards, Assistant Director of Public Health, Derbyshire County Primary Care Trust

Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene & Tropical Medicine

Dr Kenneth Robertson, Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow

Dr Catherine Swann, Associate Director, Centre for Public Health Excellence, NICE

Mrs Jean Thurston, Public contributor

Professor David Weller, Head, School of Clinical Science and Community Health, University of Edinburgh

Observers

Ms Christine McGuire, Research & Development, Department of Health

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

External Devices and Physical Therapies Panel

Members

Chair,
Dr John Pounsford,
Consultant Physician North Bristol NHS Trust

Deputy Chair,
Professor E Andrea Nelson,
Reader in Wound Healing and Director of Research, University of Leeds

Professor Bipin Bhakta,
Charterhouse Professor in Rehabilitation Medicine, University of Leeds

Mrs Penny Calder,
Public contributor

Dr Dawn Carnes,
Senior Research Fellow, Barts and the London School of Medicine and Dentistry

Dr Emma Clark,
Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol

Mrs Anthea De Barton-Watson,
Public contributor

Professor Nadine Foster,
Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University

Dr Shaheen Hamdy,
Clinical Senior Lecturer and Consultant Physician, University of Manchester

Professor Christine Norton,
Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust

Dr Lorraine Pinnington,
Associate Professor in Rehabilitation, University of Nottingham

Dr Kate Radford,
Senior Lecturer (Research), University of Central Lancashire

Mr Jim Reece,
Public contributor

Professor Maria Stokes,
Professor of Neuromusculoskeletal Rehabilitation, University of Southampton

Dr Pippa Tyrrell,
Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester

Dr Nefyn Williams,
Clinical Senior Lecturer, Cardiff University

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Interventional Procedures Panel

Members

Chair,
Professor Jonathan Michaels,
Professor of Vascular Surgery, University of Sheffield

Deputy Chair,
Mr Michael Thomas,
Consultant Colorectal Surgeon, Bristol Royal Infirmary

Mrs Isabel Boyer,
Public contributor

Mr Sankaran Chandra Sekharan,
Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust

Professor Nicholas Clarke,
Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust

Ms Leonie Cooke,
Public contributor

Mr Seumas Eckford,
Consultant in Obstetrics & Gynaecology, North Devon District Hospital

Professor Sam Eljamel,
Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee

Dr Adele Fielding,
Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School

Dr Matthew Hatton,
Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust

Dr John Holden,
General Practitioner, Garswood Surgery, Wigan

Dr Fiona Lecky,
Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust

Dr Nadim Malik,
Consultant Cardiologist/Honorary Lecturer, University of Manchester

Mr Hisham Mehanna,
Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust

Dr Jane Montgomery,
Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust

Professor Jon Moss,
Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust

Dr Simon Padley,
Consultant Radiologist, Chelsea & Westminster Hospital

Dr Ashish Paul,
Medical Director, Bedfordshire PCT

Dr Sarah Purdy,
Consultant Senior Lecturer, University of Bristol

Dr Matthew Wilson,
Consultant Anaesthetist, Sheffield Teaching Hospitals NHS Foundation Trust

Professor Yit Chiun Yang,
Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Pharmaceuticals Panel

Members

Chair,
Professor Imti Choonara,
Professor in Child Health,
University of Nottingham

Deputy Chair,
Dr Yoon K Loke,
Senior Lecturer in Clinical
Pharmacology, University of East
Anglia

Dr Martin Ashton-Key,
Medical Advisor, National
Commissioning Group, NHS
London

Dr Peter Elton,
Director of Public Health, Bury
Primary Care Trust

Dr Ben Goldacre,
Research Fellow, Epidemiology
London School of Hygiene and
Tropical Medicine

Dr James Gray,
Consultant Microbiologist,
Department of Microbiology,
Birmingham Children's Hospital
NHS Foundation Trust

Dr Jurjees Hasan,
Consultant in Medical Oncology,
The Christie, Manchester

Dr Carl Heneghan,
Deputy Director Centre for
Evidence-Based Medicine and
Clinical Lecturer, Department of
Primary Health Care, University
of Oxford

Dr Dyfrig Hughes,
Reader in Pharmacoeconomics
and Deputy Director, Centre for
Economics and Policy in Health,
IMSCaR, Bangor University

Dr Maria Kouimtzi,
Pharmacy and Informatics
Director, Global Clinical Solutions,
Wiley-Blackwell

Professor Femi Oyebode,
Consultant Psychiatrist and Head
of Department, University of
Birmingham

Dr Andrew Prentice,
Senior Lecturer and Consultant
Obstetrician and Gynaecologist,
The Rosie Hospital, University of
Cambridge

Ms Amanda Roberts,
Public contributor

Dr Gillian Shepherd,
Director, Health and Clinical
Excellence, Merck Serono Ltd

Mrs Katrina Simister,
Assistant Director New Medicines,
National Prescribing Centre,
Liverpool

Professor Donald Singer,
Professor of Clinical
Pharmacology and Therapeutics,
Clinical Sciences Research
Institute, CSB, University of
Warwick Medical School

Mr David Symes,
Public contributor

Dr Arnold Zermansky,
General Practitioner, Senior
Research Fellow, Pharmacy
Practice and Medicines
Management Group, Leeds
University

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Mr Simon Reeve,
Head of Clinical and Cost-
Effectiveness, Medicines,
Pharmacy and Industry Group,
Department of Health

Dr Heike Weber,
Programme Manager, Medical
Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

Psychological and Community Therapies Panel

Members

Chair,
Professor Scott Weich,
Professor of Psychiatry, University
of Warwick, Coventry

Deputy Chair,
Dr Howard Ring,
Consultant & University Lecturer
in Psychiatry, University of
Cambridge

Professor Jane Barlow,
Professor of Public Health in
the Early Years, Health Sciences
Research Institute, Warwick
Medical School

Dr Sabyasachi Bhaumik,
Consultant Psychiatrist,
Leicestershire Partnership NHS
Trust

Mrs Val Carlill,
Public contributor

Dr Steve Cunningham,
Consultant Respiratory
Paediatrician, Lothian Health
Board

Dr Anne Hesketh,
Senior Clinical Lecturer in Speech
and Language Therapy, University
of Manchester

Dr Peter Langdon,
Senior Clinical Lecturer, School
of Medicine, Health Policy and
Practice, University of East Anglia

Dr Yann Lefeuvre,
GP Partner, Burrage Road Surgery,
London

Dr Jeremy J Murphy,
Consultant Physician and
Cardiologist, County Durham and
Darlington Foundation Trust

Dr Richard Neal,
Clinical Senior Lecturer in General
Practice, Cardiff University

Mr John Needham,
Public contributor

Ms Mary Nettle,
Mental Health User Consultant

Professor John Potter,
Professor of Ageing and Stroke
Medicine, University of East
Anglia

Dr Greta Rait,
Senior Clinical Lecturer and
General Practitioner, University
College London

Dr Paul Ramchandani,
Senior Research Fellow/Cons.
Child Psychiatrist, University of
Oxford

Dr Karen Roberts,
Nurse/Consultant, Dunston Hill
Hospital, Tyne and Wear

Dr Karim Saad,
Consultant in Old Age Psychiatry,
Coventry and Warwickshire
Partnership Trust

Dr Lesley Stockton,
Lecturer, School of Health
Sciences, University of Liverpool

Dr Simon Wright,
GP Partner, Walkden Medical
Centre, Manchester

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

<table>
</table>

Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (www.hta.ac.uk) 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.