Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study

Helene Moustgaard,[1-4] Gemma L Clayton,[5] Hayley E Jones,[5] Isabelle Boutron,[6] Lars Jørgensen,[4] David R T Laursen,[1-4] Mette F Olsen,[4] Asger Paludan-Müller,[4] Philippe Ravaud,[6] Jelena Savović,[5,7] Jonathan A C Sterne,[5,7,8] Julian P T Higgins,[5,7] Asbjørn Hróbjartsson[1-3]

ABSTRACT

OBJECTIVES
To study the impact of blinding on estimated treatment effects, and their variation between trials; differentiating between blinding of patients, healthcare providers, and observers; detection bias and performance bias; and types of outcome (the MetaBLIND study).

DESIGN
Meta-epidemiological study.

DATA SOURCE
Cochrane Database of Systematic Reviews (2013-14).

ELIGIBILITY CRITERIA FOR SELECTING STUDIES
Meta-analyses with both blinded and non-blinded trials on any topic.

REVIEW METHODS
Blinding status was retrieved from trial publications and authors, and results retrieved automatically from the Cochrane Database of Systematic Reviews. Bayesian hierarchical models estimated the average ratio of odds ratios (ROR), and estimated the increases in heterogeneity between trials, for non-blinded trials (or of unclear status) versus blinded trials. Secondary analyses adjusted for adequacy of concealment of allocation, attrition, and trial size, and explored the association between outcome subjectivity (high, moderate, low) and average bias. An ROR lower than 1 indicated exaggerated effect estimates in trials without blinding.

RESULTS
The study included 142 meta-analyses (1153 trials). The ROR for lack of blinding of patients was 0.91 (95% credible interval 0.61 to 1.34) in 18 meta-analyses with patient reported outcomes, and 0.98 (0.69 to 1.39) in 14 meta-analyses with outcomes reported by blinded observers. The ROR for lack of blinding of healthcare providers was 1.01 (0.84 to 1.19) in 29 meta-analyses with healthcare provider decision outcomes (eg, readmissions), and 0.97 (0.64 to 1.45) in 13 meta-analyses with outcomes reported by blinded patients or observers. The ROR for lack of blinding of observers was 1.01 (0.86 to 1.18) in 46 meta-analyses with subjective observer reported outcomes, with no clear impact of degree of subjectivity. Information was insufficient to determine whether lack of blinding was associated with increased heterogeneity between trials. The ROR for trials not reported as double blind versus those that were double blind was 1.02 (0.90 to 1.13) in 74 meta-analyses.

CONCLUSION
No evidence was found for an average difference in estimated treatment effect between trials with and without blinded patients, healthcare providers, or outcome assessors. These results could reflect that blinding is less important than often believed or meta-epidemiological study limitations, such as residual confounding or imprecision. At this stage, replication of this study is suggested and blinding should remain a methodological safeguard in trials.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Blinding is an established methodological procedure in randomised clinical trials

Empirical estimates of the expected degree of bias in trials due to lack of blinding can help interpret trial results (eg, in a systematic review or clinical guideline) and plan future trials

Previous meta-epidemiological studies have reported variable estimates of the effect of blinding, with little discussion of who was blinded and the type of outcome

WHAT THIS STUDY ADDS

This large meta-epidemiological study of 142 Cochrane meta-analyses found no evidence that lack of blinding of patients, healthcare providers, or outcome assessors had an impact on effect estimates in randomised clinical trials, on average

This finding does not support the importance of blinding and is inconsistent with some previous studies; but it is consistent with several other smaller meta-epidemiological studies

The results indicate that blinding, on average, could be less important than previously believed, or could reflect limitations in the meta-epidemiological approach, such as confounding and misclassification; replication of the study is recommended and, at present, no change to methodological practice is suggested

Introduction

A randomised clinical trial is the most reliable method for assessing the effect of therapeutic interventions.[1] Results of clinical trials underpin evidence based clinical practice and decisions made by regulatory agencies, either directly or as part of a meta-analysis. However, results of randomised clinical trials might be biased[2]—for example, by systematic differences between the care provided to participants or systematic differences in the behaviour of participants, in the intervention and comparison groups (performance bias); or by systematic differences between these groups in the way in which outcomes are assessed (detection bias). Blinding (sometimes called masking) of patients, healthcare providers, and outcome assessors is intended to prevent such bias.

Blinding is used in some form in about 60% of trials.[3] However, blinding of patients and healthcare

providers is sometimes not possible owing to the type of interventions being tested (eg, psychotherapy). In other instances, blinding might not be applied owing to logistical challenges. Historically, use of placebo control interventions and blinding procedures was closely linked to early development of the randomised trial. Blinding has been an established methodological principle since around 1950.[4]

Various meta-epidemiological studies have investigated the effect of blinding on estimated intervention effects.[5][6] Such studies collate large numbers of meta-analyses of randomised trials, compare the results of blinded and non-blinded trials within meta-analyses, and then combine estimated within-meta-analysis differences across meta-analyses.[6] Estimates of the average impact of blinding have shown considerable variation between studies.[7] These studies mostly dealt with several types of bias simultaneously, and their analyses had conceptual and methodological limitations. Comparison of double blind trials with trials that are not double blinded is problematic, because the double blind concept is ambiguous.[8][9] This ambiguity is especially clear in non-pharmacological trials, and the comparison does not enable separation of performance bias and detection bias. To date, all meta-epidemiological studies of blinding have relied exclusively on information provided by trial publications, where inadequate reporting of blinding is common. Only one study took into account by whom outcomes were reported.[10]

A more comprehensive analysis of the impact of blinding in randomised trials is important. Designers of trials have to consider whether spending resources on blinding is worthwhile. Users of trial information (eg, consumers, researchers conducting systematic reviews, and guideline developers) must assess the risk of bias due to incomplete blinding.

We conducted a meta-epidemiological study to estimate the separate effects of blinding patients, healthcare providers, and outcome assessors on the results of randomised clinical trials. We also estimated the impact of different types of blinding on between-study heterogeneity.

Methods
Identification of meta-analyses for inclusion
We sought meta-analyses that included at least one trial with blinding of patients, healthcare providers, or outcome assessors (that is, observers) and at least one trial without blinding of the same groups. We refer to these as informative meta-analyses. To identify these, we screened all 1042 Cochrane reviews published or updated between 1 February 2013 and 18 February 2014 (Cochrane Database of Systematic Reviews, issue 2, 2013). We used Cochrane risk of bias tool[2] assessments to select potentially informative meta-analyses suitable for further data extraction. Specifically, we examined the first listed meta-analysis in the review's table of contents with an observer reported outcome and a difference between trials in the risk of bias score for detection bias (high v low or high v

unclear risk); and with a patient reported or healthcare provider decision outcome (outcomes determined by clinical decisions—eg, readmissions or need for surgical intervention) and a difference between trials in the risk of bias score for performance bias.

The screening process identified 395 potentially informative meta-analyses. Of these, 226 provided information on blinding of outcome assessors and 169 on blinding of patients or healthcare providers. For pragmatic reasons, we selected for further study a random subsample of 120 meta-analyses from the former set, but retained all of the latter set, giving a total of 289 potentially informative meta-analyses (full details are in the appendix).

Data retrieval and extraction
Trial publications (and any corresponding protocols/methods publications) were retrieved for each trial in each potentially informative meta-analysis. When publications could not readily be retrieved, we requested a copy from Cochrane review authors. For trials published after 1999 and where the blinding status of trial participants was unclear we contacted authors by email, asking for information on the blinding status of all groups within the trial.

We read the full text of publications in languages known to us (English, Danish, French, German, and Spanish). For publications in other languages (eg, Chinese) we based data extraction on any English language abstract, but did not attempt translation of the full text.

Data on basic trial characteristics and information on blinding status were extracted manually from trial publications. Trial results were extracted automatically from the Cochrane Database of Systematic Reviews through the Archie database interface: number of patients in intervention and control groups, for binary outcomes the number of events, and for measurement scale outcomes the means and standard deviations. We also automated extraction of the name of the Cochrane review group, and review authors' risk of bias assessments for the domains "allocation concealment" and "incomplete outcome data."

Assessment of blinding status
We assessed the blinding status of patients, healthcare providers, and outcome assessors using a modified algorithm derived from that of Akl and colleagues[11] (full details are given in the appendix). The algorithm entailed contacting trial authors (for trials published after 1999) when there was insufficient information on blinding in the trial publications. We defined blinding as a lack of awareness by patients, healthcare providers, or outcome assessor of the intervention status of individual patients throughout the trial.

We coded healthcare providers as blinded if all staff groups involved in patient treatment and care were described as "blinded" (eg, doctors and nurses, or all staff), and as non-blinded if all, or a subgroup, were described as "non-blinded" (eg, surgeons). Staff responsible for healthcare provider decision outcomes

were thus also covered by the blinding status of healthcare providers.

We differentiated between definitive information on blinding status (definitely yes/definitely no) based on explicit description or contact with trial authors, and assessments based on other information in publications (probably yes/probably no). For instance, for drug trials using a placebo control and described as "double blind" or "triple blind," patients, healthcare providers, and outcome assessors were all classified as blinded (probably yes), unless stated explicitly otherwise. For trials with no mention of "placebo," "double dummy," "double blinding," "triple blinding," "single blinding," or similar, all trial groups were classified as non-blinded (probably no), unless stated explicitly otherwise. Assessment of blinding status was made by two observers independently (AP-M, DRTL, LJ, MFO, HM, or AH), and any differences were resolved by discussion between the two. When we did not receive a reply from authors, or where we did not attempt contact, the blinding status was recorded as unclear.

When making a final determination of whether meta-analyses were informative, and for the purposes of our analyses, we compared trials that had relevant parties recorded as having "definitely no," "probably no," or "unclear" blinding with those that had relevant parties coded as "definitely yes" or "probably yes." After detailed assessment of blinding status, 189 of the 289 meta-analyses were classified as informative.

Classifications and exclusions
Classification of interventions as experimental and control was based on descriptions in the trial publications, except when the review clearly labelled the comparator as "placebo," "control," "standard care," or "treatment as usual," in which case we followed the labelling used by the review authors and classified these interventions as controls. To ensure consistent comparisons of estimated bias across meta-analyses, we excluded those meta-analyses in which intervention classifications were unclear.

Outcome measures were classified as observer reported, patient reported (via interviewer or directly recorded by patients), healthcare provider decision outcomes, or mixed (in instances where the outcome was a mixture of more than one category—eg, both patient and observer reported elements). We excluded meta-analyses of trials that did not all have the same type of outcome (eg, patient reported) unless there was an informative subset of trials with the same type of outcome.

Observer reported outcomes were subdivided into four outcomes: objective—all cause mortality, objective—other than total mortality (eg, automatised non-repeatable laboratory tests), subjective—pure observation (eg, assessment of radiographs), and subjective—interactive (eg, assessment of clinical status). Subjective observer reported outcomes were scored 1-3 according to the degree of subjectivity (that is, the extent to which determination of the outcome

depended on the judgment of the observer, with 1 indicating a low degree of subjectivity). The scoring of subjectivity was done by two observers (HM and MFO) independently and masked to any results of trials or meta-analyses, with any differences resolved by discussion. Box 1 shows examples of outcomes and subjectivity scores.

Meta-analyses were classified according to whether the outcome was measured in the trials based on an underlying hypothesis of benefit (eg, degree of pain measured based on the hypothesis that the intervention lowers pain) or of harm (eg, frequency of allergic reactions measured based on the hypothesis that the intervention could cause an increase). Classification of outcomes according to clinical area and type of experimental and comparison interventions was conducted to facilitate comparisons with an earlier meta-epidemiological study.[12] We further categorised experimental interventions as alternative/complementary or conventional medicine, to facilitate comparison with a systematic review of trials randomising patients to blinded and unblinded substudies.[13]

We excluded trials with binary outcomes, in which no or all participants had the outcome event, and trials with continuous outcomes, where the required information for calculating the standardised mean difference was missing. We also excluded trials included in more than one meta-analysis with the same outcome, if the meta-analyses were to be included in the same meta-epidemiological analysis. Such trials were removed at random until the trial occurred only within one meta-analysis. After removal of individual trials, some meta-analyses were no longer informative. The final study database contained 142 meta-analyses with a total of 1153 trials.

Data analysis
All main analyses were prespecified. In our main analyses, which included only meta-analyses with outcomes measured based on a hypothesis of benefit, we differentiated between types of bias (detection bias and performance bias) and category of person blinded (patient, healthcare provider, and outcome assessor). We performed five main analyses, quantifying the average association between estimates of treatment effect and lack of blinding:

Box 1: Examples of subjectivity scoring of trial outcomes

- Subjectivity score 1 (low degree of subjectivity): heart rate, forced expiratory volume in first second (FEV_1), cotinine saliva dipstick assay
- Subjectivity score 2 (medium degree of subjectivity): superficial surgical site infection, recurrence of varicose veins, tooth prosthesis failure
- Subjectivity score 3 (high degree of subjectivity): change in global measure of cognition, Barthel index score (of ability to perform activities of daily living), Hamilton depression scale score

- (Ia) Blinding of patients in trials with patient reported outcomes (considering a combination of detection bias and performance bias)
- (Ib) Blinding of patients in trials with blinded observer reported outcomes (considering performance bias)
- (IIa) Blinding of healthcare providers in trials with healthcare provider decision outcomes (considering a combination of detection bias and performance bias)
- (IIb) Blinding of healthcare providers in trials with blinded observers or patients assessing the outcome (considering performance bias)
- (III) Blinding of outcome assessors (that is, observers) in trials with subjective outcomes (considering detection bias).

We did not primarily focus on trials with objective outcomes, such as all cause mortality, because we did not suspect any marked effect of blinding in such trials. We conducted univariable analyses for each contrast in blinding status using all informative meta-analyses for that characteristic.

Intervention effects for binary outcomes were modelled as log odds ratios and coded such that an odds ratio of less than 1 indicated a beneficial intervention effect. For continuous outcomes, the standardised mean difference and corresponding standard error were used and coded such that a standardised mean difference of less than zero meant a beneficial intervention effect.

We quantified differences in intervention effects, comparing non-blinded trials with blinded trials of each type using ratios of odds ratios: $ROR=OR_{non-blinded}/OR_{blinded}$. Bayesian hierarchical models for meta-epidemiological research, developed by Welton and colleagues, were used to estimate the average bias associated with lack of each type of blinding (ROR), the average variability in this bias within a meta-analysis (quantified by κ, the standard deviation increase in heterogeneity between trials), and variability in average bias between meta-analyses (quantified by ϕ, the standard deviation in mean bias between meta-analyses).[14]

The model thus enabled us to explore the average degree of bias, and also whether the bias differs (eg, in direction) between meta-analyses (that is, the importance of blinding might depend on the clinical scenario) and between trials (that is, the importance of blinding might depend on factors related to the singular trial, even within similar clinical scenarios).

The analyses were carried out using Markov chain Monte Carlo simulations in WinBUGS version 1.4.3. Vague prior distributions were assumed for all parameters (see appendix for more details). We modelled continuous and binary data simultaneously, assuming a mixture of normal and binomial likelihoods but modelling the underlying bias on the same scale. This method required re-expressing standardised mean differences as odds ratios.[15] To reduce risk of spurious findings, we defined a lower threshold of at least 10 meta-analyses for conducting an analysis.

To study the impact of subjectivity scores on the average difference in intervention effect associated with blinding outcome assessors, we extended the model of Welton et al[14] to incorporate a three level categorical covariate (low v moderate v high degree of subjectivity) at the meta-analysis level.

In sensitivity analyses, we excluded trials with a classification of blinding status as "unclear" from the analyses. Secondary analyses were stratified by outcome type (eg, objective outcomes and subtypes).

Confounding by other flaws in trial design was assessed in multivariable analyses by re-running each of the five main analyses with adjustment in the model for concealment of the allocation sequence, incomplete outcome data (attrition), trial size, and blinding status of patients. The blinding status of patients was only included in the analysis of outcome assessor blinding (III). We adjusted for each of these characteristics in separate analyses. We did not include combinations of the covariates.

We also conducted post hoc subgroup analyses according to type of outcome data (continuous v binary) and type of comparator (active control v inactive control), calculated the impact of concealment of the allocation sequence on estimated treatment effects, and repeated the main analyses using an alternative label-invariant meta-epidemiological model, proposed recently by Rhodes et al.[16] This model removes the constraint that intervention effects are at least as variable among the non-blinded trials as among the blinded trials within each meta-analysis, but was not available when we wrote our protocol.

Finally, to facilitate comparison of our results with previous meta-epidemiological studies we also compared trials described by trial authors as "double blind" or "triple blind" with those not described in this way.

Patient and public involvement
Patients and members of the public were not involved in the research because it was designed to answer a methodological challenge that was not directly dependent on patient priorities, experiences, or participant preferences. The methodological expertise required to plan the study, analyse the results, and write the manuscript was dependent on specialist knowledge and we did not try to identify patients or members of the public with this training to work with.

Results
The final study database contained 142 meta-analyses with a total of 1153 trials. Figure 1 shows the flow of data through the study, from screening to final dataset. We contacted the trial author for 54 (5%) of the 1153 trials in the dataset. In 28 instances the authors replied (response rate 52%), and the fraction of trials with unclear blinding status was thereby reduced from 95/1153 (8%) to 67/1153 (6%).
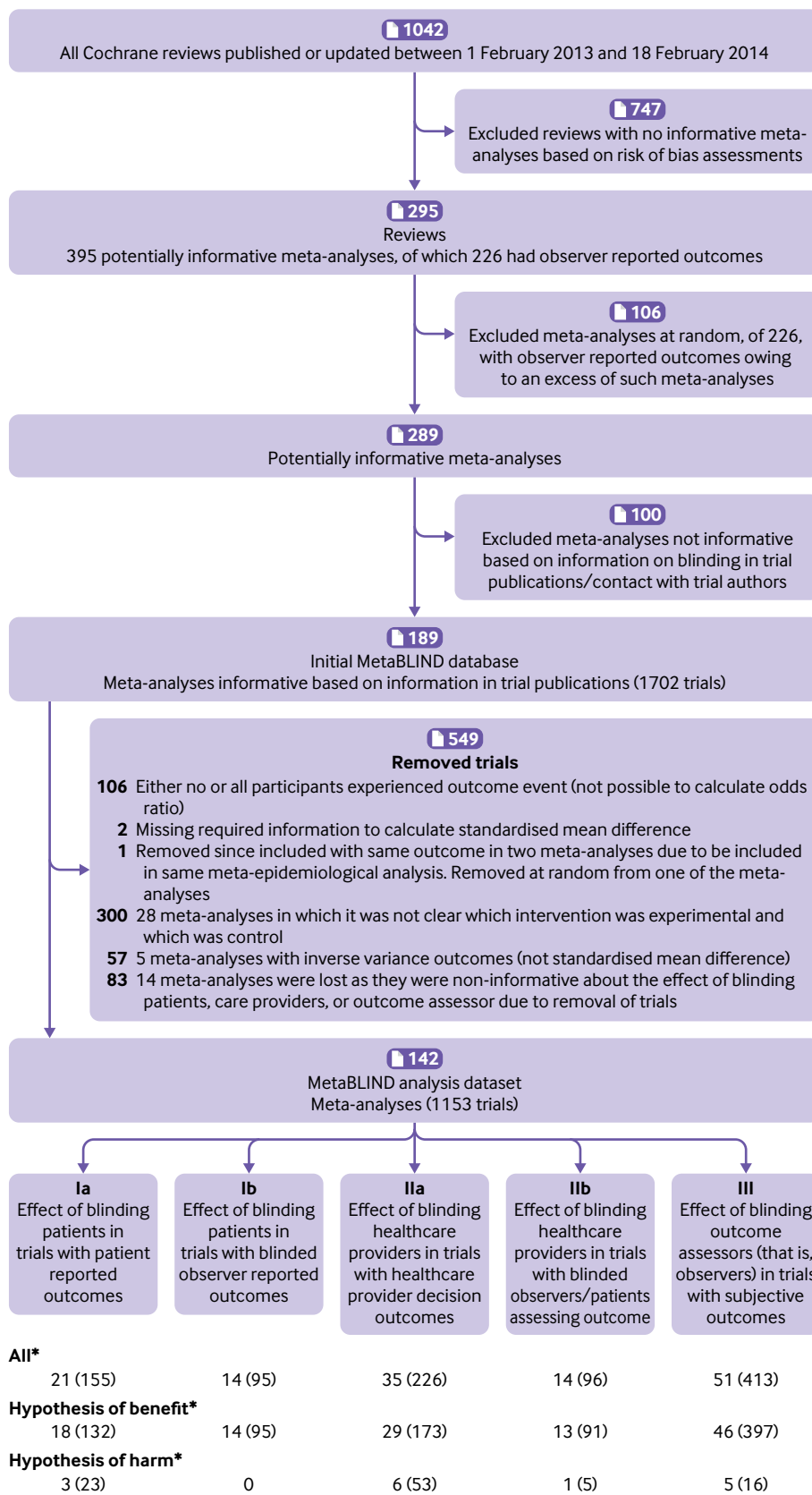
1042
All Cochrane reviews published or updated between 1 February 2013 and 18 February 2014

747
Excluded reviews with no informative meta-analyses based on risk of bias assessments

295
Reviews
395 potentially informative meta-analyses, of which 226 had observer reported outcomes

106
Excluded meta-analyses at random, of 226, with observer reported outcomes owing to an excess of such meta-analyses

289
Potentially informative meta-analyses

100
Excluded meta-analyses not informative based on information on blinding in trial publications/contact with trial authors

189
Initial MetaBLIND database
Meta-analyses informative based on information in trial publications (1702 trials)

549
Removed trials
106 Either no or all participants experienced outcome event (not possible to calculate odds ratio)
2 Missing required information to calculate standardised mean difference
1 Removed since included with same outcome in two meta-analyses due to be included in same meta-epidemiological analysis. Removed at random from one of the meta-analyses
300 28 meta-analyses in which it was not clear which intervention was experimental and which was control
57 5 meta-analyses with inverse variance outcomes (not standardised mean difference)
83 14 meta-analyses were lost as they were non-informative about the effect of blinding patients, care providers, or outcome assessor due to removal of trials

142
MetaBLIND analysis dataset
Meta-analyses (1153 trials)

	Ia Effect of blinding patients in trials with patient reported outcomes	**Ib** Effect of blinding patients in trials with blinded observer reported outcomes	**IIa** Effect of blinding healthcare providers in trials with healthcare provider decision outcomes	**IIb** Effect of blinding healthcare providers in trials with blinded observers/patients assessing outcome	**III** Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes
All*	21 (155)	14 (95)	35 (226)	14 (96)	51 (413)
Hypothesis of benefit*	18 (132)	14 (95)	29 (173)	13 (91)	46 (397)
Hypothesis of harm*	3 (23)	0	6 (53)	1 (5)	5 (16)

Fig 1 | Study flow diagram. *Meta-analyses contributing with trials that had outcome measures categorised as "mixed" (that is, it was not possible to classify them as patient reported, healthcare provider decision, or observer reported because they contained elements from more than one of these types) were not counted. Mixed outcome trials did not contribute to the main analyses

Table 1 | Characteristics of meta-analyses and trials included for the overall dataset and main analyses

Characteristics	Overall dataset		Main analyses*									
			Ia		Ib		IIa		IIb		III	
	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials
	142	1153	18	132	14	95	29	173	13	91	46	397
Outcome measures according to clinical area												
Adverse events of treatment	22 (15.5)	129 (11.2)	0	0	0	0	0	0	0	0	0	0
All cause mortality	7 (4.9)	143 (12.4)	0	0	2 (14.3)	27 (28.4)	0	0	2 (15.4)	27 (29.7)	0	0
Cause specific mortality	1 (0.7)	11 (1.0)	0	0	0	0	0	0	0	0	1 (2.2)	11 (2.8)
Clinician assessed outcomes (eg, body mass index, blood pressure, lung function, infant weight)	12 (8.5)	95 (8.2)	0	0	1 (7.1)	11 (11.6)	1 (3.4)	3 (1.7)	0	0	11 (23.9)	92 (23.2)
Composite endpoint (including mortality or major morbidity)	2 (1.4)	16 (1.4)	0	0	2 (14.3)	12 (12.6)	0	0	1 (7.7)	7 (7.7)	1 (2.2)	9 (2.3)
Global improvement	3 (2.1)	14 (1.2)	0	0	2 (14.3)	5 (5.3)	0	0	2 (15.4)	5 (5.5)	2 (4.3)	12 (3.0)
Laboratory reported outcomes (eg, blood components, tissue analysis, urinalysis)	5 (3.5)	45 (3.9)	0	0	1 (7.1)	2 (2.1)	0	0	0	0	1 (2.2)	4 (1.0)
Lifestyle outcomes (including diet, exercise, smoking)	5 (3.5)	100 (8.7)	1 (5.6)	2 (1.5)	0	0	0	0	0	0	3 (6.5)	63 (15.9)
Major morbidity event (including myocardial infarction, stroke, haemorrhage)	5 (3.5)	44 (3.8)	0	0	3 (21.4)	24 (25.3)	0	0	3 (23.1)	24 (26.4)	5 (10.9)	44 (11.1)
Mental health outcomes (including cognitive function, depression and anxiety scores)	7 (4.9)	61 (5.3)	2 (11.1)	9 (6.8)	1 (7.1)	4 (4.2)	0	0	0	0	5 (10.9)	52 (13.1)
Other outcomes (not classified elsewhere)	15 (10.6)	145 (12.6)	5 (27.8)	79 (59.8)	1 (7.1)	2 (2.1)	4 (13.8)	16 (9.2)	2 (15.4)	4 (4.4)	5 (10.9)	48 (12.1)
Pain (extent of pain a patient is experiencing)	5 (3.5)	17 (1.5)	3 (16.7)	8 (6.1)	0	0	0	0	1 (7.7)	7 (7.7)	1 (2.2)	2 (0.5)
Perinatal outcomes	5 (3.5)	34 (2.9)	0	0	0	0	1 (3.4)	2 (1.2)	0	0	1 (2.2)	9 (2.3)
Pregnancy outcomes	8 (5.6)	28 (2.4)	0	0	0	0	1 (3.4)	3 (1.7)	0	0	6 (13.0)	23 (5.8)
Quality of life (including ability to perform physical, daily, and social activities)	3 (2.1)	19 (1.6)	2 (11.1)	6 (4.5)	0	0	0	0	0	0	1 (2.2)	13 (3.3)
Radiological outcomes (including radiograph abnormalities, ultrasound, magnetic resonance imaging results)	2 (1.4)	11 (1.0)	0	0	1 (7.1)	8 (8.4)	0	0	1 (7.7)	8 (8.8)	2 (4.3)	11 (2.8)
Resource use (including cost, hospital stay duration, number of procedures)	19 (13.4)	133 (11.5)	0	0	0	0	19 (65.5)	133 (76.9)	0	0	0	0
Surgical and device related outcomes	4 (2.8)	20 (1.7)	0	0	0	0	3 (10.3)	16 (9.2)	0	0	1 (2.2)	4 (1.0)
Symptoms or signs of illness or condition	6 (4.2)	35 (3.0)	5 (27.8)	28 (21.2)	3 (21.4)	17 (17.9)	0	0	1 (7.7)	9 (9.9)	0	0
Withdrawals/dropouts/compliance	6 (4.2)	53 (4.6)	4 (22.2)	38 (28.8)	1 (7.1)	4 (4.2)	7 (24.1)	42 (24.3)	2 (15.4)	11 (12.1)	0	0
Type of experimental intervention												
Pharmacological	95 (66.9)	728 (63.1)	12 (66.7)	48 (36.4)	10 (71.4)	74 (77.9)	19 (65.5)	121 (69.9)	10 (76.9)	78 (85.7)	25 (54.3)	195 (49.1)
Surgical	3 (2.1)	12 (1.0)	1 (5.6)	4 (3.0)	0	0	0	0	0	0	1 (2.2)	4 (1.0)
Psychosocial, behavioural, or educational	17 (12.0)	204 (17.7)	1 (5.6)	42 (31.8)	3 (21.4)	17 (17.9)	3 (10.3)	10 (5.8)	1 (7.7)	2 (2.2)	9 (19.6)	101 (25.4)
Other	27 (19.0)	209 (18.1)	4 (22.2)	38 (28.8)	1 (7.1)	4 (4.2)	7 (24.1)	42 (24.3)	2 (15.4)	11 (12.1)	11 (23.9)	97 (24.4)
Field of experimental intervention												
Conventional medicine	137 (96.5)	1100 (95.4)	17 (94.4)	127 (96.2)	14 (100.0)	95 (100.0)	29 (100.0)	173 (100.0)	12 (92.3)	84 (92.3)	44 (95.7)	368 (92.7)
Alternative/complementary medicine	5 (3.5)	53 (4.6)	1 (5.6)	5 (3.8)	0	0	0	0	1 (7.7)	7 (7.7)	2 (4.3)	29 (7.3)
Type of comparison intervention												
Placebo or no treatment	57 (40.1)	442 (38.3)	8 (44.4)	36 (27.3)	1 (7.1)	11 (11.6)	12 (41.4)	47 (27.2)	2 (15.4)	16 (17.6)	17 (37.0)	160 (40.3)
Other inactive (standard care)	38 (26.8)	452 (39.2)	4 (22.2)	76 (57.6)	7 (50.0)	55 (57.9)	9 (31.0)	84 (48.6)	5 (38.5)	46 (50.5)	17 (37.0)	176 (44.3)
Active comparison	47 (33.1)	259 (22.5)	6 (33.3)	20 (15.2)	6 (42.9)	29 (30.5)	8 (27.6)	42 (24.3)	6 (46.2)	29 (31.9)	12 (26.1)	61 (15.4)
Hypothesis of benefit	114 (80.3)	971 (84.2)	18 (100.0)	132 (100.0)	14 (100.0)	95 (100.0)	29 (100.0)	173 (100.0)	13 (100.0)	91 (100.0)	46 (100.0)	397 (100.0)

Table 1 | Continued

| Characteristics | Overall dataset | | Main analyses* | | | | | | | | | |
| | | | Ia | | Ib | | IIa | | IIb | | III | |
	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials	Meta-analyses	Trials
Observer reported outcome†	68 (47.9)	640 (55.5)	0	0	14 (100.0)	95 (100.0)	0	0	10 (76.9)	73 (80.2)	46 (100.0)	397 (100.0)
All cause mortality	11 (16.2)	170 (26.6)	0	0	2 (14.3)	27 (28.4)	0	0	2 (20.0)	27 (37.0)	0	0
Other objective	4 (5.9)	39 (6.1)	0	0	1 (7.1)	2 (2.1)	0	0	0	0	0	0
Subjective	53 (77.9)	431 (67.3)	0	0	11 (78.6)	66 (69.5)	0	0	8 (80.0)	46 (63.0)	46 (100.0)	397 (100.0)
Binary or measurement scale outcome												
Binary	110 (77.5)	885 (76.8)	9 (50.0)	42 (31.8)	11 (78.6)	78 (82.1)	25 (86.2)	151 (87.3)	11 (84.6)	82 (90.1)	32 (69.6)	289 (72.8)
Continuous	31 (21.8)	265 (23.0)	8 (44.4)	87 (65.9)	3 (21.4)	17 (17.9)	4 (13.8)	22 (12.7)	2 (15.4)	9 (9.9)	14 (30.4)	108 (27.2)
Inverse variance	1 (0.7)	3 (0.3)	1 (5.6)	3 (2.3)	0	0	0	0	0	0	0	0
Year of publication of trial (median, IQR)	2003	2003 (1996-2008)	2008 (2004-11)	2005 (1997-2009)	2011 (2005-12)	2005 (1998-2009)	2009 (2005-12)	2000 (1996-2008)	2010 (2003-12)	2002 (1995-09)	2010 (2003-12)	2003 (1996-2008)
Sample size of meta-analysis/trial (median, IQR)	768 (293-2025)	106 (50-270)	706 (163-1314)	133 (78-234)	809 (173-2402)	102 (43-300)	838 (323-3103)	139 (63-483)	1085 (421-3621)	149 (61-370)	599 (289-1361)	82 (40-207)

IQR=interquartile range. Results are shown as number (%) unless stated otherwise.
*Ia=Effect of blinding patients in trials with patient reported outcomes; Ib=effect of blinding patients in trials with blinded observer reported outcomes; IIa=effect of blinding healthcare providers in trials with healthcare provider decision outcomes; IIb=effect of blinding healthcare providers in trials with blinded observers or patients assessing the outcome; III=effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes.
†Judged for the whole meta-analysis, not outcomes in individual trials.

Appendix table 1 shows the proportions of trials classified as definitely yes and probably yes.

Table 1 shows characteristics of the 142 meta-analyses and 1153 trials included in the dataset. The median year of trial publication was 2003 (interquartile range 1996-2008), and the median sample size was 768 (293-2025) patients for meta-analyses and 106 (50-270) for trials. Of the 1153 trials included in the analysis dataset, 1112 (96%) had a parallel trial design and 753 (65%) were drug trials. Full details are given in appendix table 1.

Various methodological characteristics were strongly associated across trials. For instance, trials in which the outcome assessor was blinded were more likely to have adequate allocation concealment (odds ratio 3.0, 95% confidence interval 2.2 to 4.0) and complete outcome data (2.0, 1.5 to 2.8). Trials reporting that patients were blinded were more likely to report that the outcome assessor was blinded (75.0, 38.6 to 145.8). Full details are shown in appendix tables 2 and 3. Figure 2 presents results for each of the five main analyses (Ia, Ib, IIa, IIb, III). Forest plots of the meta-analyses are shown in appendix figure 1.

For the effect of blinding patients in trials with patient reported outcomes (analysis Ia), 18 informative meta-analyses with a hypothesis of benefit contained 132 trials. Patient blinding was assessed as probably yes or definitely yes in 33 trials (25%). The average ROR was 0.91 (95% credible interval 0.61 to 1.34). The average standard deviation increase in heterogeneity between trials among non-blinded trials was very imprecisely estimated and is presented in figure 2 and appendix table 4, together with implied 95% predictive intervals for the ROR in a single trial, to facilitate interpretation. For the effect of blinding patients in trials with blinded observer reported outcomes (analysis Ib), 14 informative meta-analyses with a hypothesis of benefit contained 95 trials. Patient blinding was assessed as probably yes or definitely yes in 57 (60%) of these. The average ROR was 0.98 (95% credible interval 0.69 to 1.39).

For the effect of blinding healthcare providers in trials with healthcare provider decision outcomes (analysis IIa), 29 informative meta-analyses with a hypothesis of benefit contained 173 trials. Healthcare provider blinding was assessed as probably yes or definitely yes in 93 of these trials (54%). The average ROR was 1.01 (95% credible interval 0.84 to 1.19). For the effect of blinding healthcare providers in trials with blinded observers or patients assessing the outcome (analysis IIb), 13 informative meta-analyses with a hypothesis of benefit contained 91 trials. Healthcare provider blinding was assessed as probably yes or definitely yes in 61 trials (67%). The average ROR was 0.97 (95% credible interval 0.64 to 1.45).

For the effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes (analysis III), 46 informative meta-analyses with a hypothesis of benefit contained 397 trials. Outcome assessor blinding was assessed as probably or definitely yes in 199 of these trials (50%). The average ROR was 1.01
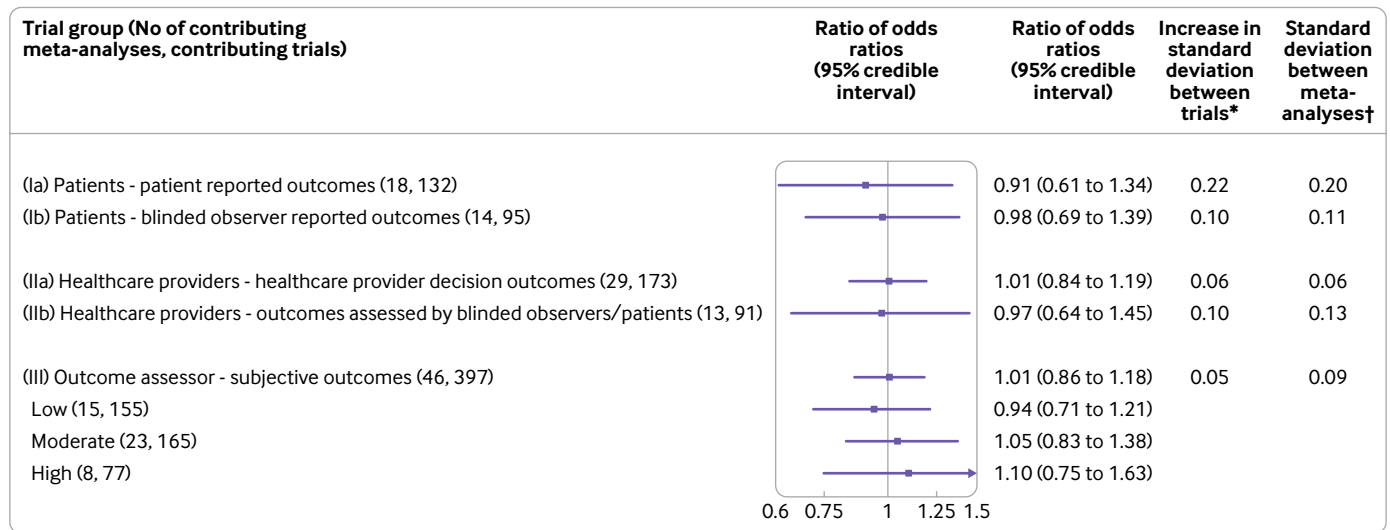
Trial group (No of contributing meta-analyses, contributing trials)	Ratio of odds ratios (95% credible interval)	Ratio of odds ratios (95% credible interval)	Increase in standard deviation between trials*	Standard deviation between meta-analyses†
(Ia) Patients - patient reported outcomes (18, 132)		0.91 (0.61 to 1.34)	0.22	0.20
(Ib) Patients - blinded observer reported outcomes (14, 95)		0.98 (0.69 to 1.39)	0.10	0.11
(IIa) Healthcare providers - healthcare provider decision outcomes (29, 173)		1.01 (0.84 to 1.19)	0.06	0.06
(IIb) Healthcare providers - outcomes assessed by blinded observers/patients (13, 91)		0.97 (0.64 to 1.45)	0.10	0.13
(III) Outcome assessor - subjective outcomes (46, 397)		1.01 (0.86 to 1.18)	0.05	0.09
Low (15, 155)		0.94 (0.71 to 1.21)		
Moderate (23, 165)		1.05 (0.83 to 1.38)		
High (8, 77)		1.10 (0.75 to 1.63)		

0.6 0.75 1 1.25 1.5

Fig 2 | Estimated ratios of odds ratios and effects on heterogeneity associated with blinding status of patients, healthcare providers, and outcome assessors. Unadjusted analyses. *Increase in standard deviation between trials: (Ia) 0.22 (95% credible interval 0.02 to 0.60), (Ib) 0.10 (0.01 to 0.30), (IIa) 0.06 (0.01 to 0.30), (IIb) 0.10 (0.01 to 0.59), (III) 0.05 (0.01 to 0.22). †Standard deviation between meta-analyses: (Ia) 0.20 (95% credible interval 0.01 to 0.74), (Ib) 0.11 (0.01 to 0.55), (IIa) 0.06 (0.01 to 0.26), (IIb) 0.13 (0.01 to 0.82), (III) 0.09 (0.01 to 0.31)

(95% credible interval 0.86 to 1.18). In the additional analysis in which we explored the impact of the level of subjectivity of the outcome, we estimated average RORs of 0.94 (0.71 to 1.21), 1.05 (0.83 to 1.38), and 1.10 (0.75 to 1.63) for outcomes with low, moderate, and high degree of subjectivity, respectively.

For each of the five main analyses, separate adjustment for concealment of the allocation sequence, attrition, and trial size did not materially change the result (table 2). Estimated increases in heterogeneity between trials and estimates of variability between meta-analyses in average bias also did not change substantially, compared with the unadjusted main analyses.

Analyses comparing trials described as "double blind" (or "triple blind") with those not so described, or with an unclear status, did not show any effect when they included meta-analyses with any type of outcome (ROR 0.99, 95% credible interval 0.86 to 1.09), nor when they included only meta-analyses with subjective observer reported outcomes and a hypothesis of benefit (1.11, 0.86 to 1.44; table 3). Exclusion of trials with an unclear blinding status from the unadjusted main analyses did not change the results substantially (table 3).

Results of secondary analyses looking separately at the effect of blinding patients, healthcare providers, or outcome assessors across different types of outcomes are shown in appendix table 5. For example, an analysis based on observer reported outcomes classified as objective also showed little evidence of an effect of outcome assessor blinding status (ROR 0.94, 95% credible interval 0.61 to 1.26; meta-analyses with a hypothesis of benefit only).

A pre-planned repetition of the main analyses based only on trials scored as definitely yes versus trials scored as definitely no proved unfeasible due to insufficient numbers of meta-analyses (appendix table 5). A post hoc analysis indicated about 10% exaggeration of the odds ratio in trials without adequate concealment of the allocation sequence (table 3). We report the results of other post hoc analyses for type of outcome (continuous v binary) and type of comparator (active control v inactive control) in table 3.

Results for the five main analyses repeated using the alternative, label-invariant, model of Rhodes et al[16] are presented in appendix table 6. The estimates of ROR and of heterogeneity between meta-analyses in bias from both models were similar. Results for heterogeneity between trials were not directly comparable to those for the main model, but indicated a possible increase in heterogeneity among blinded trials, although again the parameter estimates were very imprecise.

Discussion

We found no evidence of a difference, on average, in estimated treatment effects between randomised clinical trials with and without blinding of patients, between trials with and without blinding of healthcare providers, and between trials with and without blinding of outcome assessors. In all instances the credible intervals were wide, including both considerable difference and no difference. The same pattern was found when comparing trials that were double blind with those that were not. Our findings of an increase in heterogeneity between trials are inconclusive, owing to a lack of information.

Strengths and challenges of the study

The main strengths and originality of our study were that blinding was analysed according to the type of person blinded and due consideration given to the type of outcome. Analysis in this way allowed a separation of the two main types of blinding related bias

Table 2 | Adjusted analyses. Data are outcome measure (95% credible interval) unless stated otherwise

No of meta-analyses, trials	Adjusted for patient blinding			Adjusted for allocation concealment			Adjusted for incomplete outcome data			Adjusted for trial size		
	ROR	Φ	κ	ROR	Φ	κ	ROR	Φ	κ	ROR	Φ	κ
(Ia) Effect of blinding patients in trials with patient reported outcomes												
18, 132	NA			0.91 (0.61 to 1.35)	0.20 (0.02 to 0.74)	0.21 (0.01 to 0.61)	0.91 (0.63 to 1.31)	0.17 (0.01 to 0.70)	0.18 (0.01 to 0.60)	0.89 (0.59 to 1.29)*	0.18 (0.02 to 0.74)	0.18 (0.01 to 0.60)
(Ib) Effect of blinding patients in trials with blinded observer reported outcomes												
14, 95	NA			1.07 (0.74 to 1.56)	0.11 (0.01 to 0.57)	0.10 (0.01 to 0.57)	1.08 (0.72 to 1.58)	0.10 (0.01 to 0.52)	0.13 (0.01 to 0.54)	0.99 (0.69 to 1.39)	0.10 (0.01 to 0.54)	0.10 (0.01 to 0.57)
(IIa) Effect of blinding healthcare providers in trials with healthcare provider decision outcomes												
29, 173	NA			1.03 (0.84 to 1.23)	0.07 (0.01 to 0.29)	0.06 (0.01 to 0.29)	0.98 (0.80 to 1.17)	0.06 (0.01 to 0.28)	0.07 (0.01 to 0.28)	1 (0.83 to 1.19)	0.06 (0.01 to 0.27)	0.06 (0.01 to 0.29)
(IIb) Effect of blinding healthcare providers in trials with blinded observers/patients assessing the outcome												
13, 91	NA			1.03 (0.67 to 1.54)	0.13 (0.01 to 0.80)	0.10 (0.01 to 0.60)	1.07 (0.69 to 1.64)	0.12 (0.01 to 0.77)	0.09 (0.01 to 0.60)	0.98 (0.63 to 1.44)	0.13 (0.01 to 0.82)	0.09 (0.01 to 0.58)
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes												
46, 397	1.03 (0.87 to 1.23)	0.10 (0.01 to 0.32)	0.06 (0.01 to 0.22)	1.04 (0.89 to 1.23)	0.10 (0.01 to 0.36)	0.05 (0.01 to 0.21)	1.02 (0.87 to 1.19)	0.08 (0.01 to 0.19)	0.05 (0.01 to 0.19)	1.03 (0.88 to 1.21)	0.10 (0.01 to 0.34)	0.06 (0.01 to 0.25)

NA=not applicable; ROR=ratio of odds ratios; κ=standard deviation increase in heterogeneity between trials; Φ=standard deviation in mean bias between meta-analyses.
*One meta-analysis (three trials) was removed, which did not specify the size of the trial owing to the format given in the review.

(performance and detection bias) and enabled a comprehensive analysis that was less reliant on the way in which authors used the phrase "double blind." Also, we had a low proportion of trials with unclear blinding status, partly because we attempted to contact the trial authors. We restricted the main analyses to outcomes measured, based on a hypothesis of benefit, and ensured that interventions considered experimental in our analyses were also regarded as experimental in the individual trials.

The specificity of the comparisons limited the number of trials and meta-analyses that could be included in individual analyses, which restricted the precision of estimated differences between trials with and without the various types of blinding. We planned our sample size pragmatically, primarily based on results of comparisons within trials.[13] [17-19] Formal power calculations were published after we had planned our study.[20]

Meta-epidemiological studies are observational and so estimated effects of trial characteristics could be confounded. We adjusted for predefined variables such as allocation concealment, attrition, trial size, and blinding status of patients. Concurrent adjustment for a combination of factors was not feasible, and confounding by unknown or unmeasured factors could have affected results.

Confounding by other methodological characteristics can be expected to exaggerate the estimated effect of lack of blinding, rather than cancel it. Nevertheless, attenuation of the estimated effect of blinding by confounding cannot be ruled out. For instance, more pragmatically conducted trials within a meta-analysis (those with the broadest inclusion criteria and with least control of treatment adherence) could be less likely to have used blinding and could have resulted in less beneficial treatment effects than more explanatory trials. The consequence would be to move the estimated ROR towards 1.

Blinding could have less impact in trials comparing an experimental intervention with an active comparator (that is, not compared with placebo, no treatment, or standard care). Type of comparator, however, did not seem to affect the analysis of outcome assessor blinding, and too few informative meta-analyses precluded additional analyses. Possibly, blinding could have less impact in trials that aim to determine an intention-to-treat effect than in trials aiming to determine a per protocol effect. We did not explore whether the impact of blinding differed according to inferential goal or type of analysis.

Blinding could have been lost during the course of a trial,[21] which would tend to attenuate the apparent differences between blinded and non-blinded trials. Other factors to consider are a possibly larger impact of non-reporting bias on blinded trials, and misclassification (despite our intensive efforts to classify correctly the blinding status of patients, healthcare providers, and outcome assessors). In general, non-differential misclassification would bias our results towards no impact of lack of blinding.

9

Table 3 | Secondary analyses. Data are outcome measure (95% credible interval) unless stated otherwise

	No of meta-analyses, trials	ROR	Φ	κ
Lack of double blinding or unclear double blinding (v double blind)				
All outcomes	94, 722	0.99 (0.86 to 1.09)	0.07 (0.01 to 0.29)	0.06 (0.01 to 0.18)
Benefit	74, 583	1.02 (0.90 to 1.13)	0.06 (0.01 to 0.27)	0.07 (0.01 to 0.19)
Harms	20, 139	0.64 (0.38 to 1.04)	0.15 (0.01 to 0.89)	0.13 (0.01 to 1.23)
Observer reported outcomes: benefit	36, 374	1.04 (0.84 to 1.25)	0.14 (0.01 to 0.57)	0.08 (0.01 to 0.23)
Subjectively assessed observer reported outcomes: benefit	27, 221	1.11 (0.86 to 1.44)	0.13 (0.01 to 0.61)	0.09 (0.01 to 0.42)
Patient reported outcomes: benefit	13, 53	0.89 (0.57 to 1.40)	0.15 (0.01 to 0.83)	0.12 (0.01 to 0.88)
Healthcare provider decision outcomes: benefit	24, 147	0.98 (0.79 to 1.19)	0.07 (0.01 to 0.31)	0.07 (0.01 to 0.36)
Repeat of the main analyses excluding trials with unclear blinding status				
(Ia) Effect of blinding patients in trials with patient reported outcomes	16, 116	1.10 (0.72 to 1.69)	0.19 (0.02 to 0.76)	0.23 (0.02 to 0.61)
(Ib) Effect of blinding patients in trials with blinded observer reported outcomes	14, 94	1.00 (0.70 to 1.44)	0.11 (0.01 to 0.58)	0.10 (0.01 to 0.60)
(IIa) Effect of blinding healthcare providers in trials with healthcare provider decision outcomes	28, 160	0.97 (0.77 to 1.18)	0.08 (0.01 to 0.36)	0.07 (0.01 to 0.39)
(IIb) Effect of blinding healthcare providers in trials with blinded observers/patients assessing the outcome	13, 90	0.96 (0.64 to 1.45)	0.14 (0.01 to 0.82)	0.10 (0.01 to 0.68)
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes	43, 365	1.01 (0.85 to 1.20)	0.11 (0.01 to 0.35)	0.06 (0.01 to 0.25)
Main analysis by type of outcome				
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes: continuous*	14, 108	dSMD 0.02 (−0.22 to 0.26)	0.07 (0.01 to 0.37)	0.07 (0.01 to 0.31)
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes: binary*	32, 289	1.01 (0.85 to 1.20)	0.11 (0.01 to 0.37)	0.06 (0.01 to 0.23)
Main analyses by type of control intervention				
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes: active control*	12, 61	1.01 (0.64 to 1.55)	0.12 (0.01 to 0.70)	0.10 (0.01 to 0.56)
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes: inactive control (placebo/no treatment/standard care)*	34, 336	1.01 (0.85 to 1.21)	0.10 (0.01 to 0.36)	0.06 (0.01 to 0.23)
Inadequate (or unclear) concealment of the allocation sequence (v adequate)				
(Ia) Effect of blinding patients in trials with patient reported outcomes†	13, 116	0.95 (0.68 to 1.29)	0.11 (0.01 to 0.63)	0.10 (0.01 to 0.51)
(IIa) Effect of blinding healthcare providers in trials with healthcare provider decision outcomes†	22, 154	0.90 (0.72 to 1.12)	0.09 (0.01 to 0.35)	0.07 (0.01 to 0.32)
(III) Effect of blinding outcome assessors (that is, observers) in trials with subjective outcomes†	40, 349	0.88 (0.76 to 1.02)	0.07 (0.01 to 0.27)	0.08 (0.01 to 0.30)

ROR=ratio of odds ratios; κ=standard deviation increase in heterogeneity between trials; Φ=standard deviation in mean bias between meta-analyses; dSMD=difference in standardised mean difference.

*The prespecified minimum of 10 meta-analyses for analysis to be feasible was met only in analysis III.

†Analyses include meta-analyses from each of the datasets used in the main analyses that were informative for the impact of inadequate (or unclear) concealment of the allocation sequence. The numbers of informative meta-analyses in analyses Ib and IIb did not meet the prespecified minimum of 10 meta-analyses for analysis to be feasible.

The generalisability of our results could be affected by the sampling strategy inherent in a meta-epidemiological approach. Thus, inclusion of only meta-analyses containing both blinded and non-blinded trials excludes situations where all trials are blinded (as blinding is considered of paramount importance) or, conversely, areas where all trials tend to be non-blinded. Similarly, review authors might be more likely to include both blinded and non-blinded trials in a meta-analysis when there is no clear difference in effect estimates between the two.

Our estimation of average bias (ROR) was robust with regard to choice of statistical model.[14 16] The same applied to our analyses of heterogeneity in bias between meta-analyses. The model restriction embedded in the additive model by Welton and colleagues,[14] used for our main analyses, however, implies that between-trial heterogeneity among non-blinded trials can only increase (or remain unchanged). We reanalysed our data with an alternative model not restricted by this assumption,[16] which was not available when we planned our study. The reanalysis indicated a possible decrease in heterogeneity among non-blinded trials, although estimates were imprecise, and results were also consistent with a considerable increase in heterogeneity between trials. We interpret this result cautiously, to imply that there was insufficient information to determine whether lack of blinding was associated with increased heterogeneity between trials. Few direct comparisons have been published

between the newly developed label-invariant model[16] and the additive model[14] used in our study and in most large meta-epidemiological studies.[12 22] Analyses of the ROBES study database based on the additive model indicated an increase in heterogeneity between trials among trials with inadequate or unclear concealment of allocation, whereas the label-invariant model indicated a decrease.[16]

Other studies

Systematic reviews of meta-epidemiological studies[7 23] identified four studies (comparisons within meta-analyses) estimating the impact of blinding patients, three studies estimating the impact of blinding trial personnel, and four studies estimating the impact of blinding outcome assessors. In all instances, blinding had surprisingly little effect.[7 23] Two additional recent studies partly confirmed this pattern: an analysis of physiotherapy trials[24] found little evidence of an impact of blinding of patients or of outcome assessors, and a study of oral health trials[25] found no evidence of an impact of blinding of outcome assessors, though some evidence of a moderate effect of patient blinding.

By contrast, three systematic reviews of within-trial comparisons for 51 trials with both blinded and non-blinded outcome assessment found that blinding had a clear effect.[17-19] For example, non-blinded outcome assessors of subjective[26] outcomes exaggerated odds ratios by 36%, on average.[17] Similarly, a systematic review of 12 trials randomising patients to blinded and non-blinded substudies reported a pronounced bias due to lack of patient blinding in complementary/alternative medicine trials with patient reported outcomes, exaggerating effect sizes by 0.56 standard deviations.[13] Such comparisons within trials have no major risk of confounding. The trial design is rare, however, so to what extent the results could be generalised is not clear.

Results of meta-epidemiological studies comparing double blind trials with trials without (or unclear) double blinding have shown noticeable variation.[7] A systematic review by Page and colleagues found an overall 8% exaggeration of odds ratios in trials without double blinding (although confidence intervals overlapped no effect),[7] and an exaggeration of 23% when outcomes were subjective.[7 12]

Mechanisms and implications

Clarification of the circumstances in which blinding is important in trials, and an empirical assessment of direction and degree of bias, have important and direct implications for the design of future trials, for interpretation of trial results, and for instructions on how to assess risk of bias when conducting systematic reviews. Clarification is also pertinent to the current debate on the balance between reliability and relevance of unblinded patient reported outcome measures (PROMS),[27 28] and the relative importance of blinded explanatory trials versus unblinded pragmatic trials.[29]

Convincing theoretical reasons lead us to expect both detection and performance bias in non-blinded trials. Experimental psychology backs the notion that expectations and interest tend to shape human evaluations.[30 31] Comparisons within trials[13 17-19] provide strong evidence that in specific settings lack of blinding in trials causes considerable bias. Exactly what characterises these settings is unclear, however. We suggest that replication of our study would be valuable, as would updates of the systematic reviews of comparisons within trials, and exploration of the conditions under which blinding is more, or less, important.

Meta-epidemiological studies are often used to assess empirically dimensions of bias in randomised trials, but they could themselves be biased. For example, meta-epidemiological studies of allocation concealment have disclosed an unexpected dependence of impact on type of outcome.[12] Theoretically, impact of allocation concealment should not depend on the subjectivity of outcomes.[7 32] We suggest careful consideration of the risk of confounding and of bias, such as bias due to misclassification of methodological characteristics or due to erroneous identification of treatments as experimental and control, in meta-epidemiological studies.[33]

Blinding has been considered an essential methodological precaution in trials for decades. We did not expect to find that our study does not firmly underpin standard methodological practice. Further, our results are coherent with other meta-epidemiological studies that have reported similar results. The implication seems to be that either blinding is less important (on average) than often believed, that the meta-epidemiological approach is less reliable, or that our findings can, to some extent, be explained by lack of precision. At present, we suggest that assessors of the risk of bias in trials included in a systematic review continue to deal with the implications of lack of blinding for risk of bias, as is done in version 2 of the Cochrane risk of bias tool.[34]

In conclusion, we found no evidence of a difference, on average, in estimated treatment effect between randomised clinical trials with blinded and non-blinded patients, between trials with blinded and non-blinded healthcare providers, and between trials with blinded and non-blinded outcome assessors. The apparent lack of a major average effect of blinding on estimated treatment effects is surprising to us and is at odds with methodological standard practices. We are unclear to what extent our results show that blinding is less important than previously believed, show the limitations of the meta-epidemiological approach (eg, residual confounding), or show a lack of precision in the comparisons made. Until our study has been replicated, and we have a clearer understanding of which types of trials are susceptible to bias associated with lack of blinding, we suggest that blinding remains an important methodological safeguard in trials in which it is feasible.

AUTHOR AFFILIATIONS
[1]Centre for Evidence-Based Medicine Odense (CEBMO), Odense University Hospital, Kløvervænget 10, DK-5000 Odense C, Denmark

2Open Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark

3Department of Clinical Research, University of Southern Denmark, Odense, Denmark

4Nordic Cochrane Centre, Copenhagen, Denmark

5Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

6Cochrane France, Hôpital Hôtel-Dieu, Paris, France

7The National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West) at University Hospitals Bristol NHS Foundation Trust, Bristol, UK

8NIHR Bristol Biomedical Research Centre, University of Bristol, Bristol, UK

1   Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. JAMA 2014;312:603-5. doi:10.1001/jama.2014.8167

2   Higgins JPT, Altman DG, Gøtzsche PC, et al, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928. doi:10.1136/bmj.d5928

3   Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005;365:1159-62. doi:10.1016/S0140-6736(05)71879-1

4   Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. Bull Hist Med 1998;72:389-433. doi:10.1353/bhm.1998.0159

5   Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12. doi:10.1001/jama.1995.03520290060030

6   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. doi:10.1002/sim.1184

7   Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hróbjartsson A, Savović J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One 2016;11:e0159267. doi:10.1371/journal.pone.0159267

8   Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials 2006;3:360-5. doi:10.1177/1740774506069153

9   Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001;285:2000-3. doi:10.1001/jama.285.15.2000

10   Nüesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum 2009;61:1633-41. doi:10.1002/art.24894

11   Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. J Clin Epidemiol 2012;65:262-7. doi:10.1016/j.jclinepi.2011.04.015

12   Savović J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157:429-38. doi:10.7326/0003-4819-157-6-201209180-00537

13   Hróbjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int J Epidemiol 2014;43:1272-83. doi:10.1093/ije/dyu115

14   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. doi:10.1111/j.1467-985X.2008.00548.x

15   Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 2000;19:3127-31. doi:10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

16   Rhodes KM, Mawdsley D, Turner RM, Jones HE, Savović J, Higgins JPT. Label-invariant models for the analysis of meta-epidemiological data. Stat Med 2018;37:60-70. doi:10.1002/sim.7491

17   Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012;344:e1119. doi:10.1136/bmj.e1119

18   Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomised clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. CMAJ 2013;185:E201-11. doi:10.1503/cmaj.120744

19   Hróbjartsson A, Thomsen ASS, Emanuelsson F, et al. Observer bias in randomised clinical trials with time-to-event outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. Int J Epidemiol 2014;43:937-48. doi:10.1093/ije/dyt270

20   Giraudeau B, Higgins JPT, Tavernier E, Trinquart L. Sample size calculation for meta-epidemiological studies. Stat Med 2016;35:239-50. doi:10.1002/sim.6627

21   Bello S, Moustgaard H, Hróbjartsson A. Unreported formal assessment of unblinding occurred in 4 of 10 randomized clinical trials, unreported loss of blinding in 1 of 10 trials. J Clin Epidemiol 2017;81:42-50. doi:10.1016/j.jclinepi.2016.08.002

22   Savović J, Turner RM, Mawdsley D, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane reviews: the ROBES meta-epidemiologic study. Am J Epidemiol 2018;187:1113-22. doi:10.1093/aje/kwx344

23   Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. J Clin Epidemiol 2016;77:24-37. doi:10.1016/j.jclinepi.2016.04.005

24   Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: a meta-epidemiological study. Am J Phys Med Rehabil 2017;96:34-44. doi:10.1097/PHM.0000000000000521

25   Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, da Costa BR, Flores-Mir C. Influence of blinding on treatment effect size estimate in randomized controlled trials of oral health interventions. BMC Med Res Methodol 2018;18:42. doi:10.1186/s12874-018-0491-0

26 Moustgaard H, Bello S, Miller FG, Hróbjartsson A. Subjective and objective outcomes in randomized clinical trials: definitions differed in methods publications and were often absent from trial reports. *J Clin Epidemiol* 2014;67:1327-34. doi:10.1016/j.jclinepi.2014.06.020

27 Ghimire P, Hasegawa H, Kalyal N, Hurwitz V, Ashkan K. Patient-reported outcome measures in neurosurgery: a review of the current literature. *Neurosurgery* 2018;83:622-30. doi:10.1093/neuros/nyx547

28 Claessen FM, Mellema JJ, Stoop N, Lubberts B, Ring D, Poolman RW. Influence of priming on patient-reported outcome measures: a randomized controlled trial. *Psychosomatics* 2016;57:47-56. doi:10.1016/j.psym.2015.09.005

29 Ware JH, Hamel MB. Pragmatic trials--guides to better patient care?*N Engl J Med* 2011;364:1685-7. doi:10.1056/NEJMp1103502

30 Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998;2:175-220. doi:10.1037/1089-2680.2.2.175

31 Rosenthal R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *Am Sci* 1963;51:268-83.

32 Higgins JPT, Ramsay C, Reeves BC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. *Res Synth Methods* 2013;4:12-25. doi:10.1002/jrsm.1056

33 Moustgaard H, Jones HE, Savović J, et al. Ten questions to consider when interpreting results of a meta-epidemiological study – the MetaBLIND study as a case. *Res Synth Meth* 2019;1-15. doi:10.1002/jrsm.1392

34 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898

Supplementary information: Appendix