International Journal of Epidemiology, 2014, 1272–1283
doi: 10.1093/ije/dyu115
Advance Access Publication Date: 30 May 2014
Original article

Original article

Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies

Asbjørn Hróbjartsson,[1]* Frida Emanuelsson,[1] Ann Sofia Skou Thomsen,[1] Jørgen Hilden[2] and Stig Brorson[3]

[1]Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark, [2]Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark and [3]Department of Orthopaedic Surgery, Herlev University Hospital, Herlev, Denmark

*Corresponding author. Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark. E-mail: ah@cochrane.dk

Accepted 7 May 2014

Abstract

Background: Blinding patients in clinical trials is a key methodological procedure, but the expected degree of bias due to nonblinded patients on estimated treatment effects is unknown.

Methods: Systematic review of randomized clinical trials with one sub-study (i.e. experimental vs control) involving blinded patients and another, otherwise identical, sub-study involving nonblinded patients. Within each trial, we compared the difference in effect sizes (i.e. standardized mean differences) between the sub-studies. A difference <0 indicates that nonblinded patients generated a more optimistic effect estimate. We pooled the differences with random-effects inverse variance meta-analysis, and explored reasons for heterogeneity.

Results: Our main analysis included 12 trials (3869 patients). The average difference in effect size for patient-reported outcomes was –0.56 (95% confidence interval –0.71 to –0.41), ($I^2 = 60\%$, $P = 0.004$), i.e. nonblinded patients exaggerated the effect size by an average of 0.56 standard deviation, but with considerable variation. Two of the 12 trials also used observer-reported outcomes, showing no indication of exaggerated effects due lack of patient blinding. There was a larger effect size difference in 10 acupuncture trials [–0.63 (–0.77 to –0.49)], than in the two non-acupuncture trials [–0.17 (–0.41 to 0.07)]. Lack of patient blinding also increased attrition and use of co-interventions: ratio of control group attrition risk 1.79 (1.18 to 2.70), and ratio of control group co-intervention risk 1.55 (0.99 to 2.43).

Conclusions: This study provides empirical evidence of pronounced bias due to lack of patient blinding in complementary/alternative randomized clinical trials with patient-reported outcomes.

Key words: Bias, randomized clinical trials, patient blinding, blinding, systematic review

Key Messages

- It is reasonable to suspect bias due to nonblinded patients in randomized clinical trials, but the typical degree of bias is not known.
- We reviewed randomized clinical trials with a sub-study involving nonblinded patients and an otherwise identical sub-study involving blinded patients, providing a reliable within-trial evaluation of the bias associated with lack of patient blinding.
- In 12 trials (3869 patients) with patient-reported outcomes, nonblinded patients exaggerated the standardized mean difference by an average of 0.56 standard deviation, but with considerable variation. In trials with a true moderate effect size of –0.5, nonblinded patients thus cause an exaggeration of the estimated effect by 112%.
- The degree of bias due to nonblinded patients in two non-acupuncture trials was less than in 10 acupuncture trials, and there was no indication of bias in two trials with outcomes assessed by blinded observers.
- Generalizability of the finding is restricted by all trials investigating complementary/alternative medicine interventions.

Introduction

A longstanding methodological ideal is to keep patients in clinical trials unaware of their allocated treatment.[1] Nonblinded patients, aware of their treatment, may differ from blinded patients in how they report symptoms or in the quality of the doctor-patient relationship, inducing dissimilar rates of, for example, co-intervention, attrition and placebo effect.[2] Patient blinding is thus strongly supported by tradition and theoretical considerations.

However, the empirical foundation for this key methodological procedure is considerably less solid than often thought.[2–7] Studies of the placebo effect and of blinding of outcome assessors provide circumstantial evidence,[3–6] but the core empirical studies are large meta-epidemiological comparisons between double-blind trials and similar trials not double-blind.[7] An overview of seven meta-epidemiological studies reported an average 13% exaggeration of odds ratios in trials not double-blind, and a 22% exaggeration when outcomes were subjective.[7]

Such between-trial analyses have a considerable risk of confounding, as the blinded trials may differ from nonblinded trials for many other reasons, for example how randomization was performed, sample size or multi-centre status. This problem is augmented by the limited number of adjustments possible for suspected confounders. Thus, in meta-epidemiological studies the association between lack of double-blinding and bias is not necessarily a causal one. In addition, the term 'double-blind' is ambiguous[8] and does not enable a clear distinction between the impact of patient blinding, and blinding of care providers and outcome assessors. What most meta-epidemiological studies address is blinding in a generic sense, not specifically the blinding of patients.

It is therefore a cause of concern that the only meta-epidemiological study we have identified that explicitly compared trials that blinded patients with similar trials that did not blind patients, reported a small average difference in effect size: –0.15 (95% confidence interval –0.39 to 0.09). The confidence interval overlapped the neutral result, and the difference disappeared when the authors adjusted for concealment of allocation.[9]

It is clearly reasonable to suspect bias in randomized clinical trials with nonblinded patients. However, the average degree of bias is not known, nor is its range, variation or likely dependence on type of outcome. This is problematic because an estimate of the expected bias is important for: clinicians interpreting results from the numerous trials where patient blinding is not possible or has not been implemented; for reviewers assessing the risk of bias in trials included in a meta-analysis;[10] and for researchers who plan trials and are forced to balance the risk of bias with the practical challenges and increased cost of implementing patient blinding.

Thus, we decided to systematically review clinical trials that randomized patients to blind and nonblind sub-studies. Our primary aim was to study the impact of lack of patient blinding on estimated treatment effects in randomized clinical trials; secondary aims were to study the impact on attrition rates and use of co-interventions.

Methods

Selection of studies

We included parallel group four-armed clinical trials that randomized patients to a blinded sub-study (experimental vs control) and an otherwise identical nonblind sub-study (experimental vs control). We also included three-armed trials with experimental and no-treatment groups and a placebo group portrayed to patients as another

experimental group, i.e. patients were not informed about the possibility of a placebo intervention (nondisclosed placebo trial). This permitted the experimental group to be included both in a nonblind sub-study (experimental vs no-treatment control) and a blind sub-study (experimental vs placebo control) (Figure 1). We call these studies main trials.

In other three-armed trials, patients were informed about the possibility of a placebo intervention (disclosed placebo trial). These trials therefore lack an experimental nonblind condition and this precludes a reliable estimation of the impact of patient blinding. We included such trials in the review for comparative reasons but not in the main analyses. We call these studies auxiliary trials.

We included trials with adequately concealed allocation (e.g. central randomization or envelope method). Trials were excluded if concealment of allocation was inadequate (e.g. day of month) or unclear. Pilot studies, cross-over trials and trials with split-body design were also excluded.

Patients were regarded as blinded when this was explicitly reported or when blinding was indicated by use of a placebo treatment (and if there was no indication of unblinding of patients). We excluded trials where the placebo treatment was used to blind patients to the hypothesis of the trial and not to the treatment they received.

Patients were regarded as nonblinded when explicitly reported to be so, or when lack of blinding was indicated by use of an untreated control group. We included only trials in which the experimental and control groups were intended to receive the same basic care.

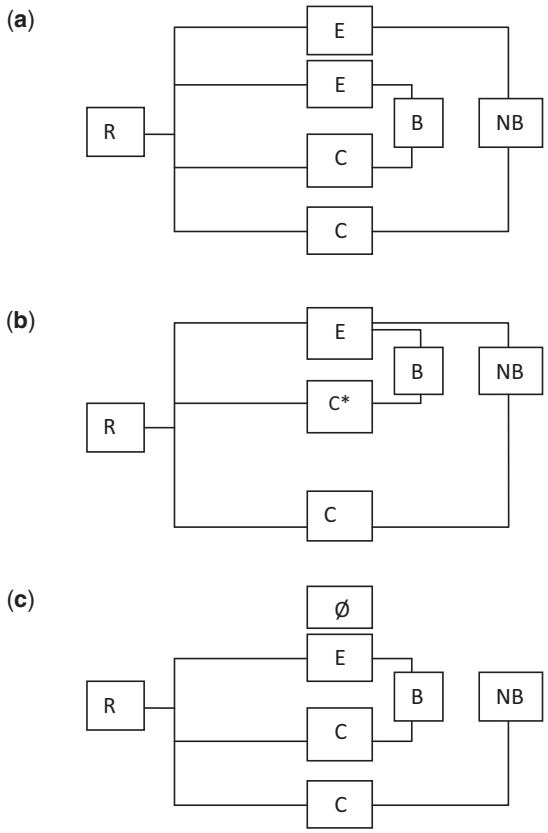
Search strategy

We searched for the four-armed trials in: Medline, EMBASE and Cochrane Methodology Register (inception to March 2013). The search for three-armed trials was based on a re-assessment of trials included in a systematic review of placebo interventions[3] and a search in the three databases mentioned above (2008–12) (Appendix 1, available as Supplementary data at *IJE* online). Reference lists to included trials were read.

Eligibility and data extraction

Reports of potentially eligible trials were read in full by one author (F.E. or A.H.), and excluded if clearly ineligible. Two authors read all other trial reports in full (F.E. or S.B. and A.H.) and decided on inclusion independently. Disagreement was resolved by discussion.

We extracted trial information using a pilot-tested data chart: clinical condition, type of intervention, number of sessions, length of treatment, type of placebo, type of no-treatment, type of co-intervention, randomization method,



Figures 1a-1c. Diagrams of the design of eligible randomized clinical trials. (**a**) Four-armed trials (main trials). (**b**) Three-armed trials with nondisclosed placebo (main trials). (**c**) Three-armed trials with disclosed placebo (auxiliary trials). R, randomized; B, blinded; NB, non-blinded; E, experimental group; C: control group; C*, placebo control presented to patients as an experimental intervention; Ø, missing non-blind experimental group. In the three-armed trials with nondisclosed placebo, patients were not informed that the trial involved a placebo but instead led to believe that two experimental interventions were compared with a no-treatment control. Such trials are equivalent to four-armed trials as the patients in the experimental intervention group are blinded with regard to the nature of the 'experimental' interventions (one is in fact a blind control), and nonblinded with regard to whether they receive an experimental intervention or no intervention (no-treatment group). In the three-armed trials with disclosed placebo, patients were informed that the trial involved a possibility of receiving a placebo, and this implies a lack of a non-blind experimental group.

number of patient randomized, content of information to patients and the blinding status of the treatment provider.

We also selected, if possible, four patient-reported measurement scale outcomes: a symptom-specific outcome [e.g. pain on a visual analogue scale (VAS)] and a global outcome (e.g. a quality of life score), timed at first post-treatment assessment and at last follow-up assessment. If possible, we also selected an observer-reported measurement scale outcome evaluated by blinded assessors, preferably timed at post-treatment. If more than one outcome in each category was eligible, we selected outcomes according to the following hierarchy: (i) primary outcome of the trial;

(ii) for symptom-specific outcomes, visual analogue scales; (iii) and for global outcomes, global assessments before quality of life measures [e.g. the short-form health survey (SF-36)]. We favoured the analysis population being as-randomized and only based on available patients (i.e. no imputations). The outcomes were selected by two authors (F.E. or S.B. and A.H.) independently and disagreements resolved by discussion.

We furthermore extracted the total number of patients in each group, the means and the standard deviations. If outcome data were missing or could not be calculated, we contacted trial authors, or used standard deviations from a similar trial with the same scale or baseline standard deviations. For trials reporting binary outcomes only, we converted odds ratios to standardized mean differences, assuming a logistic distribution,[10] and included the trials in a sensitivity analysis.

We also noted the number of drop-outs in each group (defined as number of patients unavailable for assessment) and also the number of patients in each group who used co-interventions (e.g. rescue medication), or the mean number of days (or dose or consultations) with such interventions.

Data analysis

The four-armed main trials enabled an assessment of the difference between the effect estimate based on blinded patients (blind sub-study) and the corresponding effect estimate based on nonblinded patients (nonblind sub-study).

Also three-armed main trials with nondisclosed placebo enabled an estimate of the impact of patient blinding. In such trials, all patients were aware of the distinction between the experimental intervention and the no-treatment control intervention, permitting a nonblind comparison (nonblind sub-study). Furthermore, those patients receiving an 'experimental' intervention were 'blinded' with respect to its type, which—unknown to the patients—included a placebo control, thus permitting a blind comparison between experimental and control (blind sub-study) (Figure 1).

For each of the four-armed main trials, we calculated the difference in effect sizes [i.e. standardized mean differences (SMD)], dSMD: the SMD from the sub-study based on nonblinded patients minus the SMD from the sub-study based on the blinded patients. The standardization factor used was the pooled standard deviation based on the blinded patients. For the main three-armed trials, this indirect comparison is reducible to a direct comparison of placebo and no-treatment. All within-trial comparisons were reported as dSMD, with a standard error calculated as described in Appendix 2 (available as Supplementary data at *IJE* online). A dSMD <0 indicates that nonblind patients generate more optimistic estimates of intervention effects.

We pooled the dSMDs from all trials using random-effects inverse variance meta-analyses. For our main analysis we preferred symptom-specific outcomes assessed at post-treatment; but when such data were missing we used global assessment or follow-up data as available.

We subsequently studied whether the following factors impacted on our main result: type of outcome reporting; timing of the outcome assessment; type of patient-reported outcome; type of clinical condition; type of intervention; and trial characteristics (trial design, risk of unblinding, length of trial, risk of outcome selection bias).

Our secondary study aims involved assessing the impact of lack of blinding on rates of attrition and co-intervention. We therefore calculated the attrition risk in the control groups with blinded patients and with nonblind patients, and their ratio (control group attrition risk ratio).

We furthermore calculated the co-intervention rate in the control groups with blinded patients and with nonblind patients, and their ratio (control group co-intervention risk ratio). For trials reporting the mean use of co-intervention (e.g. mean days) we calculated the standardized mean difference between the control groups with blinded and non-blinded patients.

We also meta-analysed the auxiliary three-armed trials, i.e. patients were aware of the possibility of a placebo treatment. The comparison between placebo and no-treatment for these trials will likely underestimate the impact of patient blinding, and this auxiliary analysis served only as a benchmark for the likely minimal impact of patient blinding. Ethics committee approval was not necessary for this study.

Results

We inspected 2216 database hits and 234 trial summaries[3] and read 198 full-text publications. We excluded 166 studies, in most cases because they were not randomized clinical trials with adequate concealment of allocation or were not three-armed or four-armed trials (Appendix Figure 1, available as Supplementary data at *IJE* online). Thus, we included 32 trials (7581 patients), 12 main trials[11–22] and 20 auxiliary trials.[23–42]

All 12 main trials (3869 patients) investigated complementary/alternative interventions, predominantly acupuncture (Tables 1 and 2). Patient-reported measurement scale outcomes were available for all trials; two trials also assessed observer-reported outcomes.[11,20]

Of the 20 auxiliary trials (3712 patients), four trials (150 patients) had incompletely reported outcomes.[38–41] We thus had data from 16 trials, of which four trials were only included in sensitivity analyses, one trial because of borderline eligibility[38] (Appendix 3, **available as** Supplementary data at *IJE* online) and three trials because their

Table 1. Summary characteristics of randomized clinical trials included in main analysis

Characteristics	*N*=12
Trial design	
Four-armed	2
Three-armed with nondisclosed placebo[a]	10
General characteristics	
Intervention: Non-pharmacological	11
Intervention: Pharmacological	1
Complementary-alternative intervention	12
Standard medical intervention	0
Outcome	
Patient-reported	12
Observer-reported	2
Medical specialty	
Rheumatology	6
Neurology	2
General practice, gynaecology, oncology, psychiatry	4
Trial methodology	
Random allocation sequence concealed: central	8
Random allocation sequence concealed: other than central, e.g. envelopes	4
Treatment provider blind	1
Available protocol shows no sign of outcome reporting bias	7
Number of patients per trial (median)[b]	303 (248–381)

[a]Patients were falsely informed that the trial compared two types of experimental interventions against a no-treatment control condition, i.e. the patients in experimental group were both blinded (regarding experimental vs nondisclosed placebo) and non-blinded (regarding experimental vs no-treatment control).

[b]Interquartile range.

binary outcomes had been converted to standardized mean differences.[35–37] Thus, 12 auxiliary trials were analysed in detail (Appendix Table 2, available as Supplementary data at *IJE* online).

Main trials: primary results

The differences between effect sizes based on the 12 sub-studies with blinded patients and those of the corresponding sub-studies with nonblinded patients ranged from –0.12 to –1.06. The average effect size difference was –0.56 (95% confidence interval –0.71 to –0.41), with notable heterogeneity ($I^2 = 60\%$, $P = 0.004$) (Figure 2). Thus, in trials with patient-reported outcomes, effect sizes based on nonblinded patients were exaggerated by an average of 0.56 standard deviation.

For this main analysis, symptom-specific post-treatment outcomes were used in 10 trials.[11,13–19,21,22] a global post-treatment outcome in one trial[12] and a symptom-specific follow-up outcome in one trial.[20] However, there was no clear tendency for results to differ depending on type of patient-reported outcome (Table 3).

Two main trials had outcomes assessed by blinded observers.[11,20] One trial of *Echinacea purpurea* for the common cold reported a median (but not mean) increase in IL-8 with no clear difference in effect between the sub-study with 350 blinded patients and the sub-study with 360 nonblinded patients.[11] Another trial[20] investigated the effect of acupuncture on mean TUG score (time to up and go in seconds) in 560 patients with osteoarthritis, difference in effect size –0.02 (–0.22 to –0.18) (Appendix 4, available as Supplementary data at *IJE* online).

Differences in effect sizes for subgroups are presented in Table 3. Two main trials were four-armed (1128 patients), and investigated the effect of distant healing for chronic fatigue syndrome[12] and *Echinacea* for the common cold.[11] The average difference in effect size was –0.17 (–0.41 to 0.07), ($I^2 = 0\%$, $P = 0.69$) (Table 3). Ten main trials were three-armed (2741 patients) and investigated the effect of acupuncture on pain, nausea or anxiety.[13–22] The average difference in effect size was –0.63 (–0.77 to –0.49), ($I^2 = 43\%$, $P = 0.07$).

Main trials: secondary results

Two main trials of short duration (2 weeks or less) had no drop-outs.[18,22] The average risk of patient attrition in the nonblinded control groups of the 10 trials of medium to long duration (more than 2 weeks) was 7% (4% to 11%). The similar average blinded control group attrition risk was 4% (2% to 6%). The average ratio of control group attrition risk was 1.79 (1.18 to 2.70), i.e. the risk of patient drop-out was 79% higher in the nonblind control group as compared with the blinded control group.

Based on the two main trials that provided binary co-intervention data,[15,16] more patients in the nonblind control group used co-intervention than in the blind control group, ratio of control group co-intervention risk 1.55 (0.99 to 2.43). Based on five main trials that provided measurement scale co-intervention outcomes,[13,14,17–19] the average use of co-intervention was also higher in the nonblinded control group, standardized mean difference –0.29 (–0.13 to –0.45)

Auxiliary trials

We analysed in detail 12 auxiliary trials (1536 patients) without a nonblind experimental condition (Appendix Table 2, available as Supplementary data at *IJE* online), difference in effect size –0.26 (–0.37 to –0.16).

One of the 12 trials used blinded observers to assess outcomes, difference in effect sizes –0.06 (–0.57 to 0.45).[25] Two

Table 2. Individual characteristics of main trials

Trial	Condition (N[1])	Trial design	Nonblind patients		Blind patients		Patient-reported outcome	Timing		Analyses	
			Experimental	Control	Experimental	Control	Symptom-specific Global/Quality of life[2]	PT[3]	FU[4]	As R[5]	Missing data[6]
Walach 2008	Chronic fatigue syndrome (409)	Four-armed	Distant healing	No distant healing	Distant healing	No distant healing	Symptom-specific: na[7]	m. 6[8]	na	yes	LOCF[9]
Barrett 2010	Common cold (719)	Four-armed	Echinacea capsule	No capsule	Echinacea capsule	Placebo capsule (inert ingredients, identical coating)	AUC:[10] WURSS-21 score (1-7) x time[11]	w. 2[12]	na	yes	Imputation
Brinkhaus 2006	Low back pain (301)	Three-armed with non-disclosed placebo[13]	na[13]	No acupuncture	Acupuncture[13] (MS,[14] Qi[15])	Placebo acupuncture (superficial needling in non-AP,[16] not MS or Qi)	Low Back Pain (VAS 0-100)[17]	w. 8	na	yes	Available data
Cherkin 2009	Low back pain (638)	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (MS, Qi)	Placebo acupuncture (non-penetrative "needling" in relevant AP, not MS or Qi)	Pain bothersomeness (0-10)	w. 8	na	yes	Available data
							Global: na				
Foster 2007	OA (352)[18]	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (Qi)	Placebo acupuncture (non-penetrative "needling" in relevant AP, not Qi)	WOMAC pain subscale (1-10)[19]	w. 6	w. 52	yes	Available data
Linde 2005	Migraine (302)	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (MS, Qi)	Placebo acupuncture (superficial needling in non-AP, not MS or Qi)	Days with moderate/severe headache	w. 12	na	yes	Available data
Melchart 2005	Tension headache (270)	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (MS, Qi)	Placebo acupuncture (superficial needling in non-AP, not MS or Qi)	Days with headache	w. 12	na	yes	Available data
Shen 2000	Emesis (104)	Three-armed with nondisclosed placebo	na	No acupuncture	Electroacupuncture (Qi)	Placebo acupuncture (needling in irrelevant AP, not MS or Qi, no current)	Number of emesis periods per person	d. 5[20]	d. 14	yes	no missing data
Witt 2005	OA (300)	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (MS, Qi)	Placebo acupuncture (superficial needling in non-AP, not MS, not Qi)	WOMAC index (range na)	w. 8	na	yes	Available data
Suarez 2010	OA (560)	Three-armed with nondisclosed placebo	na	No acupuncture	Electroacupuncture (Qi ns)	Placebo acupuncture (superficial needling in irrelevant AP, minimal current)	J-MAP score (1-7)[21]	na	m. 3	yes	LOCF
Liu 2011	Dysmenorrhoea (200)	Three-armed with nondisclosed placebo	na	No acupuncture	Electroacupuncture (Qi)	Placebo acupuncture (needling in irrelevant AP or in non-AP, Qi, current)	Pain VAS (0-100)	h. 1[22]	na	yes	Available data
Karst 2007	Anxiety (67)	Three-armed with nondisclosed placebo	na	No acupuncture	Acupuncture (Qi ns)	Placebo acupuncture (non-penetrative "needling" in irrelevant AP)	Anxiety, STAI X1 (20-80)[23]	h. 0.5	na	na	Available data

[1]Number of randomised patients may differ from number of analysed patients; [2]first line for symptom-specific outcomes (e.g. pain) and second line for global outcomes (e.g. overall improvement); [3]post-treatment; [4]follow-up; [5]patients analysed as randomised; [6]missing data approach; [7]not assessable/no data; [8]month; [9]last observation carried forwards; [10]area under curve; [11]Wisconsin upper respiratory symptom survey (short version); [12]week; [13]These trials involved only three groups: experimental, placebo, and no-intervention, but as patients were unaware of the placebo group (regarded it a second experimental group) the experimental intervention group was both blinded (as compared with placebo) and nonblinded (as compared with no-treatment); [14]manual stimulation; [15]de Qi is a feeling of numbness, distension, or electrical tingling at the needling site; [16]acupuncture point; [17]visual analogue scale; [18]osteoarthritis; [19]The Western Ontario and McMaster Universities Arthritis Index; [20]days; [21]Joint-specific multidimensional assessment of pain, [22]hour; [23]State-Trait Anxiety Inventory.

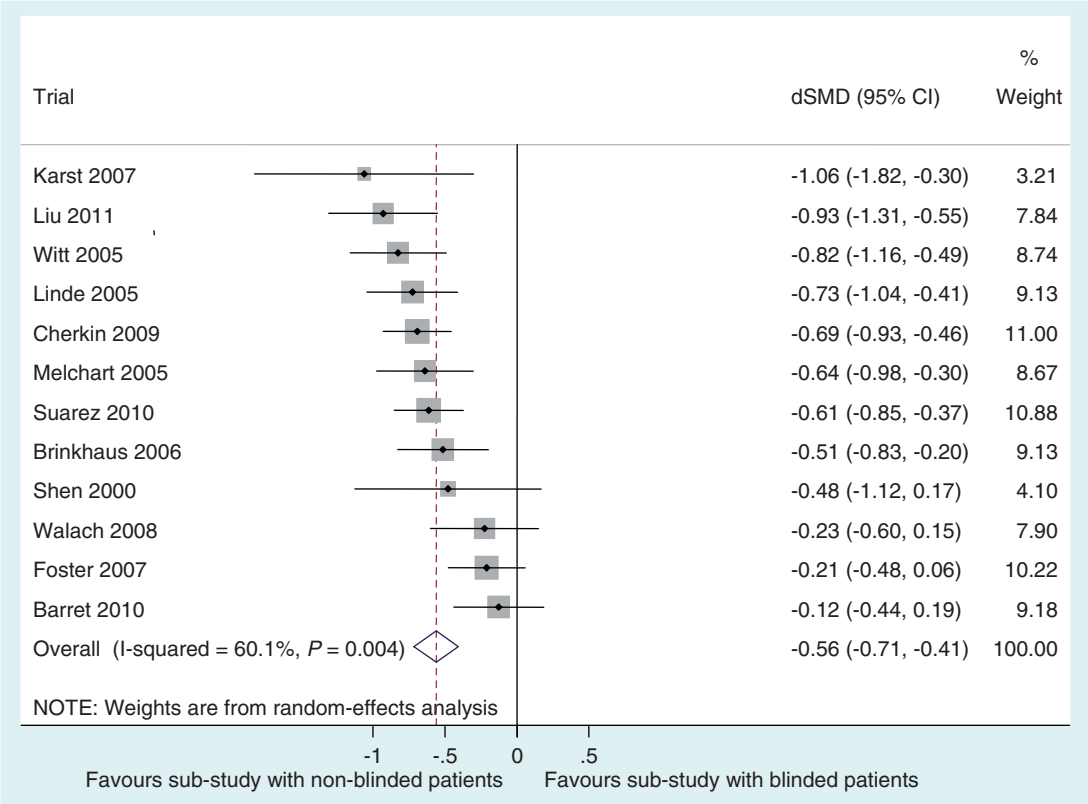


Figure 2. Meta-analysis of randomized clinical trials with sub-studies of blinded and nonblinded patients. dSMD: the difference between SMD (standardized mean difference, or effect size) based on blinded patients and the corresponding SMD based on nonblinded patients (95% confidence interval). A dSMD <0 indicates than nonblinded patients generate more optimistic estimates of intervention effect.

additional trials with binary outcomes, that we converted to effect sizes, also used blinded observers, difference in effect sizes 0.02 (–0.24 to 0.29)[35] and 0.02 (–0.54 to 0.58).[37]

Differences in effect sizes for subgroups and for sensitivity analyses are reported in Appendix Table 3 (available as Supplementary data at *IJE* online). There were consistent differences in effect sizes for eight complementary/alternative trials, –0.29 (–0.44 to –0.14) and for four standard medical trials, –0.25 (–0.44 to –0.06).

Discussion

Based on 12 randomized clinical trials, lack of blinding of patients exaggerated effect sizes by an average of 0.56 (0.71 to 0.41) standard deviation when outcomes were patient-reported, but with considerable variation. No notable effect exaggeration was found in two trials when outcomes were assessed by blinded observers. All trials investigated complementary/alternative medical interventions, predominantly acupuncture. Average attrition rates were modest, but 79% (18% to 270%) higher in the nonblind control group. Co-intervention rates showed a similar pattern.

Study strengths and limitations

The strength of our study is the comparability of the involved groups and the distinct focus on patient blinding. The randomization of patients to blind and nonblind conditions avoided confounding by differences in patient prognosis and outcomes or unknown factors—a challenge for meta-epidemiological studies.[7,9] In addition, selection bias was minimized as all trials had adequate allocation concealment and were generally large. In general, the trials were well planned and performed. Thus, observed differences in estimated effects, attrition rates and co-intervention rates between the groups with blinded and nonblinded patients were, with a very high probability, related to their blinding status.

Assuming that the bias due to nonblinded patients in the included trials is comparable to that of ordinary two-armed trials, our results may be extrapolated to acupuncture trials, and to complementary/alternative trials. We included no main standard medical trials, but analysis of the auxiliary trials found no clear difference between the degree of bias in eight complementary/alternative trials and in the four standard medical trials, indicating no marked difference in bias due to trial type. However,

Table 3. Difference in effect size according to subgroups and sensitivity analyses: main trials

Comparisons	N[a]	I^2 (P)	dSMD (95% CI)[b]
Main analysis[c]	12	60% (0.004)	−0.56 (−0.71 to −0.41)
Patient-reported outcome			
Post-treatment, symptom-specific	10	63% (0.004)	−0.59 (−0.77 to −0.41)
Post-treatment, global	7	71% (0.02)	−0.35 (−0.58 to −0.13)
Follow-up, symptom-specific	4	75% (0.008)	−0.25 (−0.54 to 0.03)
Follow-up, global	2	39% (0.001)	−0.18 (−0.42 to 0.05)
Observer-reported outcome, blind assessor[d]	1	NA	−0.02 (−0.22 to 0.18)
Type of clinical condition[e]			
Osteoarthritis	3	78% (0.01)	−0.54 (−0.88 to −0.20)
Migraine/tension headache	2	0% (0.71)	−0.69 (−0.92 to −0.46)
Low back pain	2	0% (0.37)	−0.63 (−0.82 to −0.44)
Other conditions	5	72% (0.007)	−0.52 (−0.89 to −0.14)
Type of intervention			
Complementary/alternative intervention	12	60% (0.004)	−0.56 (−0.71 to −0.41)
Standard medical intervention	0	NA	NA
Pharmacological intervention: no	11	49% (0.03)	−0.60 (−0.75 to −0.46)
Pharmacological intervention: yes	1	NA	−0.12 (−0.44 to 0.19)
Patient-provider interaction: more than 2 sessions	9	45% (0.07)	−0.62 (−0.76 to 0.48)
Patient-provider interaction: 1–2 sessions	3	60% (0.08)	−0.34 (−0.74 to 0.06)
Intervention by blind treatment provider	1	NA	−0.12 (−0.44 to 0.19)
Intervention by nonblind treatment provider	11	49% (0.03)	−0.60 (−0.75 to −0.46)
Trial characteristics			
Four-armed trials/non-acupuncture	2	0% (0.69)	−0.17 (−0.41 to 0.07)
Three-armed trials/acupuncture	10	43% (0.07)	−0.63 (−0.77 to −0.49)
Risk of patient unblinding: yes	10	43% (0.07)	−0.63 (−0.77 to −0.49)
Risk of patient unblinding: no	2	0% (0.69)	−0.17 (−0.41 to 0.07)
Short duration (0–2 weeks)[f]	2	24% (0.25)	−0.73 (−1.30 to −0.17)
Medium to long duration (>2 weeks)	10	65% (0.002)	−0.55 (−0.71 to −0.38)
Protocol: shows no outcome selection bias[f]	7	65% (0.01)	−0.45 (−0.65 to −0.26)
Protocol: unavailable/sign of outcome selection bias	5	0% (0.46)	−0.74 (−0.90 to −0.57)

[a]Number of trials; [b]pooled difference between standardized mean difference (SMD) based on blind patients and the corresponding SMD based on nonblind patients (95% confidence interval)—a dSMD <0 indicates than nonblind patients generate more optimistic estimates of intervention effect; [c]one trial with follow-up outcome, one trial with global outcome, otherwise all were post-treatment symptom-specific outcomes; [d]two main trials had observer-reported outcomes, but one reported only medians; [e]all conditions investigated in two trials or more are listed; [f]post-hoc analyses, the other subgroup and sensitivity analyses were predefined.

confidence intervals were wide, and extrapolation of our results beyond complementary/alternative trials may be problematic.

Approximately 60% (i.e. the I^2 value) of the variation in the difference of effect sizes was caused by clinical or other systematic discrepancies between the trials. This degree of heterogeneity is usually considered moderate to substantial,[10] and unconditional pooling of results may be debated. We noted that the heterogeneity was quantitative (i.e. differed in degree, not in direction), pooled the results with a random-effects model and explored reasons for the heterogeneity. We found that acupuncture trials may have a more pronounced degree of bias than non-acupuncture trials, but it was not possible to clarify in more detail whether this difference was caused by type of intervention (non-pharmacological or pharmacological), trial design (three- or four-arm trial) or variation in patient expectations.

Other studies

Nüesch and colleagues[9] conducted a meta-epidemiological study of 16 meta-analyses of osteoarthritis trials, and reported only a small difference between trials with blinded patients and similar trials with nonblinded patients, difference in effect size –0.15 (–0.39 to 0.09). However, they also found larger effect size difference for non-pharmacological trials (–0.71), and for trials of

complementary/alternative interventions (–0.48), similar to what we found. The discrepancy between our pooled results may thus be caused both by a different sample of trials or by confounding in the study by Nüesch and colleagues.

On average, nonblinded observers in randomized clinical trials with subjective outcomes exaggerate effect sizes by –0.23 (–0.40 to –0.06).[5] In trials with patient-reported outcomes, patients can be considered introspective outcome assessors of their private symptoms. Thus, it appears that trials with nonblinded patients assessing their own symptoms may cause a higher degree of bias compared with trials with nonblinded observers assessing subjective outcomes.

Meta-epidemiological studies have analysed other types of bias, for example single-centre status or patient exclusions, and have reported effect size difference that were much smaller than what we have found for lack of patient blinding: single-centre status –0.08 (–0.17 to 0.01)[43] after adjustment for sample size, and patient exclusions –0.13 (–0.29 to 0.04).[44]

Patient blinding and trial aims

The aim of many randomized clinical trials is predominantly explanatory,[45] i.e. to assess the effect of a clearly defined causal component of an intervention (for example a molecule with possible antibiotic properties) under ideal conditions. Such trials are often described as assessing 'efficacy'[46] or 'specific effects'[47] and to prioritize 'internal validity'.[48] In a trial with a predominantly explanatory aim, patient blinding adheres to the experimental ideal of keeping all factors equal except the causal component under investigation. For the purpose of our investigation, we regarded all sub-studies as having a mainly explanatory aim, and our results are therefore mainly relevant for such trials.

However, the aim of some randomized trials is mainly pragmatic, i.e. to evaluate the effect of an intervention package in a setting as close to the practical clinical context as possible. Such trials are often described as assessing 'effectiveness',[46] may include 'non-specific effects'[47] and prioritize 'external validity'.[48] In such trials, patient blinding is sometimes implemented to protect against response bias when outcomes are patient-reported.[49] However, patient blinding may interfere with the overall pragmatic aim of the study. For example, attrition rates and co-intervention rates may simultaneously be considered causes of bias and consequences of the intervention. Similarly, placebo effects, or 'unspecific effects',[3,47] may be characterized as bias or clinical effects depending on perspective.

None of the trials included in our review explicitly characterized the aim of their sub-studies as either explanatory or pragmatic. Many trials will not fall neatly into a binary classification of 'pragmatic' vs 'explanatory'. The concepts are probably best considered as describing a 'multidimensional continuum',[50,51] and often a single trial will aspire to incorporate both aspects. Typically, the challenge for trials with a predominantly explanatory aim is whether the effect estimates based on ideal conditions and selective patients can be reproduced in real-life clinical settings. However, the challenge for trials with a predominantly pragmatic aim is to disentangle bias from true effects. A tension between the diverse roles of patient blinding in trials with mainly explanatory and pragmatic aims is especially clear in acupuncture trials.[52,53]

Bias mechanisms

In trials with patient-reported outcomes, bias due to the lack of patient blinding is mainly caused by a combination of response bias (i.e. a tendency for patients to report symptoms in a way they think is expected), placebo effect, differential attrition and differential co-intervention. Further causes of concern in some trials is treatment switches, i.e. patients allocated to the experimental intervention received the control intervention (or vice versa), or contamination, i.e. patients received unintended experimental treatment.

The average impact of these bias mechanisms will likely depend on an interaction between the type of outcome, and patients' and treatment providers' predispositions which may vary considerably. Patients' effect expectations may have been low in the two non-acupuncture trials.[11,12] Previous trials of *Echinacea* for the common cold had found no clear effects,[11] and distant healing has an implausible mechanism of action.[12] Furthermore, in both trials patient-provider interaction was limited. The typical acupuncture trial, however, involves repetitive close encounters with nonblinded treatment providers, the penetration of the body surface by needles, an ancient non-biological background theory which is presented as compatible with modern biology via the gate theory.[54] These procedures may induce high effect expectations and induce a higher degree of bias.

Attrition rates and co-intervention rates were modest, though higher in the nonblinded control group. These two types of bias mechanisms will likely partly cancel each other out, with an unpredictable net impact. It appears that in our sample of trials, the main mechanism of bias was the combined influence of response bias and placebo effect.[55]

In trials where outcomes are assessed by blinded observers, response bias is not an issue and placebo effects are probably absent or considerably reduced, on average. We had access to results from five such trials (1239 patients),[10,19,24,34,36] of which two were main trials,[10,19]

providing no indication of bias due to nonblinded patients. However, to exclude the existence of a moderate degree of bias many more trials are needed.

Interpretation of effect size and degree of bias

Conventionally, standardized mean differences of –0.2 are considered small effects, –0.5 medium effects and –0.8 large effects.[56] By these standards, our overall result represents a medium effect. However, a degree of bias should not be interpreted in the same way as a treatment effect. The appropriate question is how much bias is to be expected when patients are nonblinded, not whether that degree of bias represents a clinically worthwhile effect had it been a treatment effect.

An average exaggeration of 0.56 standard deviation is: pronounced for interventions with medium 'true' effect sizes of –0.5, i.e. exaggeration of effect by 112%; and sizeable for the minority of interventions with large effect sizes of –0.8, i.e. exaggeration of effect by 70%. Thus, we interpret the degree of average bias as considerable.

Implications

We suggest that patients should be blinded whenever possible, and thus to devote considerable resources to developing and assessing patient blinding procedures, especially in trials with patient-reported outcomes.

Blinding patients is likely also important for trials with observer-reported outcomes, but the empirical evidence is less clear and the degree of bias may be less pronounced. Patient blinding is sometimes not possible, for example in trials of exercise, surgery or psychotherapy, or not regarded appropriate in trials with a predominantly pragmatic aim. Our results provide a tentative empirical framework for interpretation of results from such trials.

We emphasize that it is important not only to note the average degree of bias we have found, but also its range and variation. We are therefore sceptical of devising a uniform correction factor for compensating the expected bias in trials without patient blinding.

In conclusion, based on within-trial comparisons, this study provides empirical evidence of pronounced bias due to lack of patient blinding in complementary/alternative randomized clinical trials with patient-reported outcomes.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

The study was partially funded by the Danish Council for Independent Research: Medical Sciences. The funder had no influence on study design, the collection, analysis and interpretation of data, the writing of the article and the decision to submit it for publication.

Acknowledgements

Thanks to Peter C. Gøtzsche and Iain Chalmers for valuable comments on previous versions of the manuscript. A.H. conceived the idea and design, organized the study and wrote the first draft of the manuscript. F.E. and A.H. developed the search strategy. F.E., A.S.S.T., S.B. and A.H. did the data collection. A.H. and J.H. did the statistical analyses. All authors discussed the result and commented on the manuscript. All authors had full access to all the data in the study. A.H. takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of interest: F.E. is an employee of Novo Nordisk A/S. However, no Novo Nordisk products were involved in the study as the author's contribution to the manuscript was finalized before the current position. The other authors have declared no conflicts of interest.

References

1. Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;**72**:389–433.
2. Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: imposed impartiality. *Clin Pharmacol Ther* 2011;**90**:732–36.
3. Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2010;**1**:CD003974.
4. Hróbjartsson A, Thomsen ASS, Emanuelsson F *et al*. Observer bias in randomized clinical trials with binary outcomes: systematic review of trials with both blinded and nonblinded outcome assessors. *BMJ* 2012;**344**:e1119.
5. Hróbjartsson A, Thomsen ASS, Emanuelsson F *et al*. Observer bias in randomized clinical trials with measurement scale outcomes: systematic review of trials with both blinded and nonblinded outcome assessors. *CMAJ* 2013;**185**:E201–11.
6. Hróbjartsson A, Emanuelsson F, Skou Thomsen AS *et al*. Bias due to lack of patient blinding in clinical trials: A systematic review of trials randomizing patients to blind and non-blind substudies. *Int J Epidemiol* 2014;**43**:1272–83.
7. 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* 201;**157**:429–58.
8. Devereaux PJ, Manns BJ, Ghali WA *et al*. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;**285**:2000–03.
9. 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.
10. Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Updated February 2011. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org (9 May 2014, date last accessed).

11. Barrett B, Brown R, Rakel D *et al*. Echinacea for treating the common cold: a randomized trial. *Ann Intern Med* 2010;**153**: 769–77.
12. Walach H, Bosch H, Lewith G *et al*. Effectiveness of distant healing for patients with chronic fatigue syndrome: a randomized controlled partially blinded trial (EUHEALS). *Psychother Psychosom* 2008;**77**:158–66.
13. Witt C, Brinkhaus B, Jena S *et al*. Acupuncture in patients with osteoarthritis of the knee: a randomized trial. *Lancet* 2005; **366**:136–43.
14. Melchart D, Streng A, Hoppe A *et al*. Acupuncture in patients with tension-type headache: randomized controlled trial. *BMJ* 2005;**331**:376–82.
15. Cherkin DC, Sherman KJ, Avins AL *et al*. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* 2009;**169**: 858–66.
16. Foster NE, Thomas E, Barlas P *et al*. Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomized controlled trial. *BMJ* 2007;**335**:436.
17. Linde K, Streng A, Jürgens S *et al*. Acupuncture for patients with migraine: a randomized controlled trial. *JAMA* 2005;**293**: 2118–25.
18. Shen J, Wenger N, Glaspy J *et al*. Electroacupuncture for control of myeloablative chemotherapy-induced emesis: A randomized controlled trial. *JAMA* 2000;**284**:2755–61.
19. Brinkhaus B, Witt CM, Jena S *et al*. Acupuncture in patients with chronic low back pain: a randomized controlled trial. *Arch Intern Med* 2006;**166**:450–57.
20. Suarez-Almazor ME, Looney C, Liu Y *et al*. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. *Arthritis Care Res (Hoboken)* 2010;**62**:1229–36.
21. Liu CZ, Xie JP, Wang LP *et al*. Immediate analgesia effect of single point acupuncture in primary dysmenorrhea: a randomized controlled trial. *Pain Med* 2011;**12**:300–07.
22. Karst M, Winterhalter M, Münte S *et al*. Auricular acupuncture for dental anxiety: a randomized controlled trial. *Anesth Analg* 2007;**104**:295–300.
23. Moffett JA, Richardson PH, Frost H, Osborn A. A placebo controlled double blind trial to evaluate the effectiveness of pulsed short wave therapy for osteoarthritic hip and knee pain. *Pain* 1996;**67**:121–27.
24. O'Brien B, Relyea MJ, Taerum T. Efficacy of P6 acupressure in the treatment of nausea and vomiting during pregnancy. *Am J Obstet Gynecol* 1996;**174**:708–15.
25. Petri N, Svanholt P, Solow B, Wildschiødtz G, Winkel P. Mandibular advancement appliance for obstructive sleep apnoea: results of a randomized placebo controlled trial using parallel group design. *J Sleep Res* 2008;**17**:221–29.
26. Steinsbekk A, Bentzen N, Fønnebø V, Lewith GT. Randomized controlled trials on treatment by homeopaths and self-treatment with homeopathic medicines: design and protocol. *J Altern Complement Med* 2004;**10**:1027–32.
27. Conn IG, Marshall AH, Yadav SN, Daly JC, Jaffer M. Transcutaneous electrical nerve stimulation following appendicectomy: the placebo effect. *Ann R Coll Surg Engl* 1986;**68**: 191–92.
28. Lander J, Fowler-Kerry S. TENS for children's procedural pain. *Pain* 1993;**52**:209–16.
29. Licciardone JC, Stoll ST, Fulda KG *et al*. Osteopathic manipulative treatment for chronic low back pain: a randomized controlled trial. *Spine* 2003;**28**:1355–62.
30. Pach D, Brinkhaus B, Roll S *et al*. Efficacy of injections with Disci/Rhus toxicodendron compositum for chronic low back pain – a randomized placebo-controlled trial. *PLoS One* 2011;**6**:e26166.
31. Schwartz MP, Wellink H, Gooszen HG, Conchillo JM, Samsom M, Smout AJ. Endoscopic gastroplication for the treatment of gastro-oesophageal reflux disease: a randomized, sham-controlled trial. *Gut* 2007;**56**:20–28.
32. Wise RA, Bartlett SJ, Brown ED *et al*. Randomized trial of the effect of drug presentation on asthma outcomes: the American Lung Association Asthma Clinical Research Centers. *J Allergy Clin Immunol* 2009;**124**:436–44, 444e1–8.
33. Werntoft E, Dykes AK. Effect of acupressure on nausea and vomiting during pregnancy. A randomized, placebo-controlled, pilot study. *J Reprod Med* 2001;**46**:835–39.
34. Vas J, Aranda JM, Modesto M *et al*. Acupuncture in patients with acute low back pain: a multicentre randomized controlled clinical trial. *Pain* 2012;**153**:1883–89.
35. Vas J, Aranda-Regules JM, Modesto M *et al*. Using moxibustion in primary healthcare to correct non-vertex presentation: a multicentre randomized controlled trial. *Acupunct Med* 2013;**31**:31–38.
36. Alkaissi A, Evertsson K, Johnsson VA, Ofenbartl L, Kalman S. P6 acupressure may relieve nausea and vomiting after gynecological surgery: an effectiveness study in 410 women. *Can J Anaesth* 2002;**49**:1034–39.
37. Van Der Heijden GJ, Leffers P, Wolters PJ *et al*. No effect of bipolar interferential electrotherapy and pulsed ultrasound for soft tissue shoulder disorders: a randomized controlled trial. *Ann Rheum Dis* 1999;**58**:530–40.
38. Scharf HP, Mansmann U, Streitberger K *et al*. Acupuncture and knee osteoarthritis: a three-armed randomized trial. *Ann Intern Med* 2006;**145**:12–20.
39. Yan T, Hui-Chan CW, Li LS. Functional electrical stimulation improves motor recovery of the lower extremity and walking ability of subjects with first acute stroke: a randomized placebo-controlled trial. *Stroke* 2005;**36**:80–85.
40. Alford JW, Fadale PD. Evaluation of postoperative bupivacaine infusion for pain management after anterior cruciate ligament reconstruction. *Arthroscopy* 2003;**19**:855–61.
41. Cesarone MR, Incandela L, Belcaro G *et al*. Two-week topical treatment with essaven gel patients with diabetic microangiopathy. A placebo-controlled, randomized study. *Angiology* 2001;**52**(**Suppl 3**):S43–48.
42. De Sanctis MT, Incandela L, Belcaro G, Cesarone MR. Topical treatment of venous microangiopathy in patients with venous ulceration with Essaven gel. A placebo controlled, randomized study. *Angiology* 2001;**52**(**Suppl 3**):S29–34.
43. Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ* 2012;**344**:e813.
44. Nüesch E, Trelle S, Reichenbach S *et al*. The effects of excluding patients from the analysis in randomized controlled trials: meta-epidemiological study. *BMJ* 2009;**339**:b3244.

45. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009;**62**:499–505.
46. Price D, Hillyer EV, van der Molen T. Efficacy versus effectiveness trials: informing guidelines for asthma management. *Curr Opin Allergy Clin Immunol* 2013;**13**:50–57.
47. Hróbjartsson A. What are the main methodological problems in the estimation of placebo effects? *J Clin Epidemiol* 2002;**55:** 430–35.
48. Tugwell P, Knottnerus A, Idzerda L. Comparative effectiveness reviews need to pay as much attention to external validity as to internal validity risks of bias. *J Clin Epidemiol* 2011;**64**:1163–65.
49. Fransen GA, van Marrewijk CJ, Mujakovic S. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. *BMC Med Res Methodol* 2007;**7**:16.
50. Thorpe KE, Zwarenstein M, Oxman AD *et al*. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ* 2009;**180**:E47–57.
51. Thorpe KE, Oxman AD, Treweek S, Furberg K. Pragmatic trials are randomized and may use a placebo. *J Clin Epidemiol* 2009; **62**:464–75.
52. Vickers Vickers AJ, Cronin AM, Maschino AC; Acupuncture Trialists' Collaboration. Acupuncture for chronic pain: individual patient data meta-analysis. *Arch Intern Med* 2012;**172**: 1444–53.
53. Madsen MV, Gøtzsche PC, Hróbjartsson A. Acupuncture treatment for pain: systematic review of randomized clinical trials with acupuncture, placebo acupuncture, and no acupuncture groups. *BMJ* 2009;**338**:a3115.
54. Melzack R. Acupuncture and pain mechanisms. *Anaesthetist* 1976;**25**:204–07.
55. Hróbjartsson A, Kaptchuk TJ, Miller FG. Placebo effect studies are susceptible to response bias and to other types of biases. *J Clin Epidemiol* 2011;**64**:1223–29.
56. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, 1977.