Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2007;36:654–663
doi:10.1093/ije/dym020

Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding

A Hróbjartsson,[1]* E Forfang,[1] MT Haahr,[1] B Als-Nielsen[1] and S Brorson[2]

Accepted 30 January 2007

Background Blinding can reduce bias in randomized clinical trials, but blinding procedures may be unsuccessful. Our aim was to assess how often randomized clinical trials test the success of blinding, the methods involved and how often blinding is reported as being successful.

Methods We analysed a random sample of blinded randomized clinical trials indexed in the The Cochrane Central Register of Controlled Trials and published in 2001. We identified 1599 blinded trials, and noted if they had conducted any test for the success of blinding. We also selected 200 trials randomly that did not report any such test, and sent a questionnaire to the corresponding authors asking them if they had conducted any tests.

Results Thirty-one out of 1599 trials (2%) reported tests for the success of blinding. Test methods varied, and reporting was generally incomplete. Blinding was considered successful in 14 out of the 31 trials (45%) and unclear in 10 (32%). Of the seven trials (23%) reporting unsuccessful blinding the risk of a biased trial result was either not addressed or was discounted in six cases. We received 130 questionnaires from trial authors (65%) of which 15 (12%) informed that they had conducted, but not published, tests.

Conclusions Blinding is rarely tested. Test methods vary, and the reporting of tests, and test results, is incomplete. There is a considerable methodological uncertainty how best to assess blinding, and an urgent need for improved methodology and improved reporting.

Keywords Double-Blind Method, Statistical Data Interpretation, Randomized Controlled Trials/*methods/standards

Background

Blinding can reduce bias in randomized clinical trials. A meta-analysis of studies of bias found that trials described as 'double blind' in trial reports found 14% lower treatment effects, on average, than similar trials not described as 'double blind'.[1]

Bias may occur due to inadequate blinding of several key trial persons. Blinded trial participants may report symptoms differently than unblinded ones,[2] or differ in their tendency to drop out of the trial, or seek non-protocolized treatment. Blinded health care providers may also differ from unblinded ones in their degree of attention to patients[3] or in their use of alternative forms of care. Similarly, blinded data collectors,[4,5] end-point committees and data analysts may differ from unblinded ones in their handling of data.

It is unclear how often key trial persons intended to be blind (e.g. patients, healthcare providers, data collectors, end-point committees or data analysts) are unblinded during a trial. Critical appraisal of the risk of unblinding is often difficult because the reporting of the blinding status of key trial persons, and of the blinding procedures, is frequently incomplete or missing,[6,7] despite the CONSORT statement's suggestion to report these issues.[8]

[1] The Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark

[2] Department of Orthopedic Surgery, Amager University Hospital, Italiensvej 1, DK-2300 Copenhagen S, Denmark

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

Another possible way of assessing whether blinding has been successful is to ask key trial persons to guess patients' treatments and compare the answers with the actual treatments. Previous analyses of such trials indicated that unblinding could be common but were based on small samples of trials published in a few selected journals,[9–11] or were restricted to trials that had reported testing the success of blinding in the titles or abstracts.[11] Therefore, the results may not be generalizable.

Our objective was to assess how often randomized clinical trials test the success of blinding, which methods are used, and how often blinding is reported to be successful.

Methods

We developed a database of blinded randomized trials published in 2001 by searching The Cochrane Central Register of Controlled Trials 2003, Issue 1, using the search terms: 'random* and (blind* or mask* or placebo* or sham* or mock* or fake* or dummy* or vehicle*)'. We identified 5079 references.

We arbitrarily aimed at identifying at least 25 trials that tested the success of blinding. Based on computer-generated random numbers we initially selected 500 of the 5079 references, read the corresponding trial reports and assessed the likely number of trial reports needed. Secondarily, we randomly selected a further 1750 references, and retrieved the corresponding publications.

From this initial sample of 2250 references, we excluded 632 non-randomized or non-blinded trials, laboratory experiments, studies on healthy volunteers or health economics analyses, as well as 19 non-English articles that we could not read. The database thus consisted of 1599 blinded randomized clinical trials published in 2001.

All trial reports were read by one reviewer to identify trials that reported tests for blinding. To make sure we had not overlooked any eligible trials 400 randomly selected trials (25%) were read a second time by a different reviewer. No overlooked trials were found.

From the database of 1599 trials we defined two groups of trials. The first group consisted of all trials that reported tests of the success of blinding in the journal article. The second group consisted of 200 of the remaining trials that did not describe such tests, sampled by a computer-generated list of random numbers. No formal sample size calculation was made.

From both groups of trials, we extracted the design of the trial, the experimental and control interventions, indicators of trial quality (generation of allocation sequence, concealment of allocation, double blinding, intention-to-treat analysis, funding source and size of trial),[1,12,13] the clinical speciality involved, the key trial persons intended to be blinded, and the time of unblinding. Two authors independently extracted the data on pretested forms and any disagreements were solved by discussion.

We considered a trial blinded if the trial report used the term 'blinded' (or similar, e.g. 'masked') to describe the intention of keeping at least one key trial person ignorant of the allocated treatment of patients. We pragmatically considered a trial as 'double blind' if described by that term in the trial report; and similarly we regarded the result as analysed by 'intention to treat' if the trial report used that term. We defined adequate generation of the allocation sequence as computer-generated random numbers, random numbers lists, flip of coin, drawing of cards or lots or comparable methods of stochastic generation. Adequate concealment of the allocation sequence was defined as: central randomization (including pharmacy controlled), numbered or coded vehicles, sealed or opaque envelopes or comparable methods of concealment.

From the first group of trials that published tests of the success of blinding we furthermore extracted the following data: the type of key trial persons tested, the number of tested persons, the response categories, the results, the use of any statistical test and the trial report conclusion. We classified the success of blinding according to the conclusion of the authors of the trial reports.

We sent an e-mail to the primary authors of the 200 trials in the second group that did not test the success of blinding. The e-mail provided a link to a web-based questionnaire. To those without an identifiable e-mail address we posted a printed version. We asked the authors whether they had formally tested the success of blinding, despite not having reported such a test. If a test had been conducted we subsequently asked the author how the test was performed and what the result was. Reminders were sent out after 2, 4 and 6 weeks. The questionnaire included additional items on blinding of key trial persons. Results based on these questions is reported elsewhere.[7]

Proportions were compared using Fisher's exact test, and medians using Mann–Whitney's test. *P*-values are two-sided.

Results

Thirty-one out of the 1599 trial publications (2%; 95% confidence intervals: 1.3 to 2.7%) reported tests for the success of blinding (Table 1, Appendix,[14–44]).

Trials that publish tests for the success of blinding ($n = 31$)

Blinding was concluded successful in 14 of the 31 trials (45%; 27 to 64%), unsuccessful in seven trials (23%; 10 to 41%), whereas 10 trials (32%; 17 to 51%) had unclear or no conclusions.

The majority of trials was of parallel group design, investigated a drug and used a placebo control (Table 1). Twenty-five trials (81%) were described as 'double blind', 11 (35%) had adequate concealment of allocation and six (19%) conducted intention-to-treat analyses. The four most frequent medical specialties or research areas represented were psychiatry, public health, anaesthesiology and complementary alternative medicine (Table 1). The median number of included patients per trial was 56 (interquartile range 40–83). The median number of persons tested for the success of blinding per trial was 40 (interquartile range 20–64).

The trial reports described explicitly the blinding of patients in 26 trials (84%), and the blinding of data

Table 1 Characteristics of included trials

	Published tests[a] $N=31$, n (%)	No published tests[b] $N=200$, n (%)	P-value
Design			
Parallel group	27 (87)	171 (86)	1.00
Drug intervention	23 (74)	171 (86)	0.12
Placebo control	24 (77)	135 (68)	0.30
Trial quality dimensions			
Double blind	25 (81)	156 (78)	0.82
Adequate generation of allocation sequence	10 (32)	48 (24)	0.37
Adequate allocation concealment	11 (35)	37 (19)	0.05
Intention to treat analysis	6 (19)	57 (29)	0.39
Public funding	10 (32)	41 (21)	0.16
Trial size (median, interquartile range)	56 (40–83)	58 (30–149)	0.38
Clinical specialty[c]			
Psychiatry	12 (39)	24 (12)	0.001
Public health	3 (10)	1 (0.5)	0.008
Anaesthesiology	3 (10)	29 (15)	0.59
Complementary-alternative medicine	3 (10)	8 (4)	0.17
Allergology/pulmonology	1 (3)	18 (9)	0.21
Cardiology	0 (0)	22 (11)	0.05

[a] From our main database of 1599 trials, 31 (2%) published tests for the success of blinding.

[b] From our main database, we randomly selected 200 trials not publishing tests for the success of blinding; N: total number of trials in each group; n (%): number and proportion of trials.

[c] The four most frequent specialties in each cohort listed.

Table 2 Proportions of combinations of blinded patients/data collectors, tested patients/data collectors and type of primary outcome[a]

Blinded person	Tested person	Type of primary outcome	$N=31$, n (%)
Data coll	Data coll	Observer rep, involving patients' cooperation	4 (13)
Data coll	Data coll	Observer rep, not involving patients' cooperation	1 (3)
Pt + data coll	Pt + data coll	Observer rep, involving patients' cooperation	2 (6)
Pt + data coll	Pt	Observer rep, involving patients' cooperation	1 (3)
Pt + data coll	Pt	Patient rep	1 (3)
Pt	Pt	Observer rep, involving patients' cooperation	8 (26)
Pt	Pt	Observer rep, not involving patients' cooperation	1 (3)
Pt	Pt	Patient rep	13 (42)

[a] Data coll: data collector; Pt: patient; rep: reported. Other key trial persons than patients and data collectors were disregarded (e.g. healthcare providers). By 'tested' is implied tests for the success of blinding. 'Observer rep, not involving patients' cooperation' describe outcomes that could have been measured on unconscious patients, e.g. laboratory values; 'Observer rep, involving patients' cooperation' describe outcomes that implies that patients cooperate, e.g. scores of depression.

collectors in nine trials (29%) (Table 2 and Appendix). Four of these trials described the blinding of both patients and data collectors. In 15 trials (48%) one or more key trial person, explicitly described as blinded in the trial report, was not tested for the loss of blinding. The primary outcome was patient reported (e.g. pain) in 14 trials (45%), and observer reported (e.g. assessment of sedation success) in 17 trials (55%) (Table 2 and Appendix). Patients' cooperation was involved in 15 of the 17 trials with observer reported outcomes (e.g. raters scoring potentially depressed patients on Hamilton Rating Scale for Depression).

The methods used to test the success of blinding varied considerably. Most trials tested only the blinding of patients and asked them to guess between experimental or control treatment, without the option of a 'don't know' category (Table 3). Twenty-three trials tested patients only (74%). Four trials tested data collectors only (13%), three trials both patients and data collectors (10%) and one trial healthcare providers and data collectors (3%).

In 30 trials (97%) the types of key trial persons who were tested were directly involved in the reporting of the primary outcome. In 14 of these 30 trials patients were tested in trials with patient reported outcomes. In 15 other trials with observer

Table 3 Characteristics and conclusions of published tests of the success of blinding[a]

	$N = 31$, n (%)
Key trial persons tested in each trial	
Patients only	23 (74)
Data collectors only	4 (13)
Others or combinations	4 (13)
Test persons' guessing options	
Control or experimental	18 (58)
Control or experimental or do not know	9 (29)
Control or two types of experimental	4 (13)
Summary of test results	
Full data for 2×2 or 2×3 tables	7 (23)
Some data, not enough for full tables	18 (58)
No summary data or unclear	6 (19)
Statistical analysis	
Chi-square or Fisher's exact test	8 (26)
Trend test	1 (3)
Logistic regression	1 (3)
Kappa-value	1 (3)
Unclear	2 (6)
Not reported	18 (58)
Conclusion	
Successful blinding	14 (45%)
Unsuccessful blinding	7 (23%)
Conclusion unclear or not reported	10 (32%)

[a] From our main database of 1599 trials, 31 (2%) published tests for the success of blinding; *N*: total number of trials; *n* (%): number and proportion of trials.

Table 4 Characteristics and conclusions of unpublished tests of the success of blinding[a]

	$N = 15$, n (%)
Key trial persons tested in each trial	
Patients	7 (47)
Observers	3 (20)
Treatment providers	1 (7)
Others (e.g. patients and observers)	4 (27)
Test persons' guessing options	
Control or experimental	5 (33)
Control or experimental or do not know	6 (40)
No information	4 (27)
Statistical analysis	
Chi-square or Fisher's exact test	5 (33)
No formal analysis	9 (60)
No information	1 (7)
Conclusion	
Successful blinding	11 (73%)
Unsuccessful blinding[b]	2 (13%)
Conclusion unclear or not reported	2 (13%)

[a] From our main database, we randomly selected 200 trials not publishing tests for the success of blinding. Of the 130 responding trial authors 15 had conducted, but not published, tests. *N*: total number of trials; *n* (%): number and proportion of trials.

[b] He questionnaire wording was partial loss of blinding.

reported outcomes that involved patients' cooperation, either patients or data collectors were tested. Only in two such trials were both patients and data collectors tested (Appendix and Table 2).

Twenty-eight trials (90%) tested persons only at the end of treatment whereas three trials (10%) also tested at various times during the trial. Three trials (10%) rated the confidence of tested persons in their guesses or asked for reasons for guesses. No trials tested whether the experimental and control treatments appeared identical before the proper trial started by e.g. asking volunteers to identify any variation in colour, taste and texture.

The results were analysed with Chi-square or Fisher's exact tests in eight trials (26%), and three trials (10%) used other tests (Table 3). One trial (3%) used the test as a measure for effect and not an indication for possible bias as it assumed that blinding could only be compromised by the effect of treatment.[43]

The reporting was generally incomplete. Ten trial reports (32%) contained no clear conclusion concerning the result of the test. In 24 trials (77%) it was impossible to reconstruct a table that related key trial persons' treatment guesses to actual treatment allocation. In most cases only the proportion of correct, or incorrect, guesses was reported. In 18 trials (58%) no statistical analysis was presented, and in an additional two trials (6%) the result of the analysis was unclear (Table 3).

The seven trials with tests that indicated unsuccessful blinding differed in their way of handling this information. One trial report emphasized the risk of bias when discussing its result.[44] A second report emphasized that loss of blinding did not necessarily imply bias.[19] Five other trial reports made no clear statement concerning the risk of bias,[29,34,36,37,43] but interpreted the loss of blinding as caused by an effect of the treatment,[29,34,37] by lack of effect in the placebo group,[36] or by side effects to the treatment.[43] Four of the seven trials with unsuccessful blinding concluded that the experimental treatment had an effect on at least one outcome without discussing the possibility of bias.[29,34,36,37] Thus, in six out of seven cases the risk of bias was either not addressed or was discounted.

Trials that did not publish tests for the success of blinding ($n = 200$)

The majority of trials were of parallel group design, investigated a drug and used a placebo control (Table 1). There were 156 (78%) trials described as 'double blind', 37 (19%) had adequate concealment of allocation and 57 (29%) conducted intention-to-treat analyses. The median number of patients was 58 (interquartile range 30–149). The four most frequent clinical specialties represented were anaesthesiology, psychiatry, cardiology and allergology/pulmonology (Table 1).

We received 130 of the 200 questionnaires (response rate 65%). Of the 130 trial report authors, 15 (12%) responded that they had formally tested the success of blinding without

reporting this in the journal article (Table 4). Eleven of these 15 responders (73%) reported success of blinding, two (13%) reported 'partial loss of blinding', and two did not report the result (13%). Nine of the 15 trial authors (60%) informed that the data on the success of blinding had been assessed without a formal statistical test.

Trials that publish tests vs other trials

Among the trials that published tests for the success of blinding there was a higher proportion of psychiatry and public health trials compared with the other trials (Table 1). There were no marked differences with respect to other trial characteristics or indicators of methodological quality (Table 1). These findings were not sensitive to the exclusion of 15 trials with unpublished tests (data not shown). Thus, there was no clear tendency for trials that test the success of blinding to be of marked higher methodological quality than trials in general.

Discussion

Few trialists published tests of the success of blinding. There was considerable variation in how tests were conducted, and a high prevalence of incompletely, and unreported, tests. Less than half of the trial reports concluded that blinding had been successful.

Strength and weaknesses of the study

Our study is based on a large, recent and representative sample of trials across a broad range of study designs, specialities and journals. We identified a moderately sized group of trials that reported a test of blinding. Furthermore, we studied the proportion of unreported tests by contacting authors of trial reports that did not describe tests of blinding.

However, the reporting on tests was often incomplete. Roughly three out of four trials did not report full data for a table relating test persons' guesses to actual treatment. Furthermore, one of eight trial authors declared unpublished tests in their questionnaire. The decision whether to publish the result, and with what degree of detail, involve considerable subjectivity. There is a strong incentive for successful blinding, as this strengthens the validity of the trial. Possible selective reporting[49] could have resulted in underreporting of unsuccessful blinding. Thus, our review reflects the results of published tests, but not necessarily that of conducted tests.

Comparison with other studies

Fergusson and colleagues analysed 191 placebo-controlled trials published in top journals.[10] They identified 15 trials (8%) that tested the success of blinding, of which nine trials (60%) reported unsuccessful blinding. Boutron and colleagues analysed 127 trial reports in top journals, and searched medical data bases for trial reports that mentioned tests in the title or abstract.[11] They identified 13 trials that tested blinding published in top journals (8%), and 82 in total. Of the 54 trials assessing the success of blinding of patients, 22 trials (41%) reported unsuccessful blinding.

However, results from these studies may not be generalizable because a small sample of placebo-controlled trials could differ from trials in general. Furthermore, authors that describe a test in the title or abstract may do so partly depending on the result of the test. Thus, our study adds considerably improved generalizability, because we sampled randomized clinical trials directly from a database, and comprehensiveness because the number of included trials is larger, and because we also analysed a group of trials that did not report tests.

Methodological uncertainty of how best to evaluate the success of blinding

Both inadequate randomization and blinding is associated with bias.[1] Randomization is often evaluated by comparing the baseline characteristics of patients.[45] Blinding can be evaluated by an informal assessment of the reported blinding practices, but this approach is subjective, and highly dependent on adequate reporting of the blinding procedures, which is very uncommon.[6,7]

The alternative approach is to assess the success of blinding with a formal test. However, there is no methodological consensus of whether such tests are appropriate, and if so, how, and at what time, they should be conducted.[10,46] A positive test conducted during, or after the end of, the trial, cannot be interpreted as clear indication of bias, as unblinding may be caused by a true effect, and lack of blinding may not cause bias (e.g. if the outcome is mortality). Sackett suggests conducting a test on volunteers before the proper trial starts,[46] but this procedure does not evaluate whether blinding was maintained during the trial. There is also uncertainty of the best way to ask patients, e.g. whether questions should include a do not know category. Furthermore, the statistical problems of how to analyse data from such tests have only recently been analysed, and the suggested solutions need further validation.[47,48,50,51]

This methodological uncertainty is reflected in our findings. Most trials tested only one out of several intended blinded key trial persons. The selection criteria for who to test was often unclear, e.g. patients, and not data collectors, were tested in 10 trials with observer reported primary outcomes. Roughly half of the trials did not include a 'don't know' option when testing, and few trials had follow up questions that could elucidate reasons for guesses. The statistical analyses were in many cases unclear, and in some cases absent. Furthermore, six out of seven trial reports describing a test that indicated loss of blinding either disregarded, or ignored, the risk of bias due to unblinding. All in all, in many trials that test the success of blinding these problems preclude a definite conclusion on the blinding status of key trial persons.

Implications

In most trials critical appraisal of the success of blinding is not meaningful because of grossly incomplete reporting of both the blinding procedures,[6,7] and any test for the success of blinding, and because of the methodological uncertainty concerning such tests. As lack of blinding is associated with bias[1–4] this calls for urgent improvement.

First, trialists, peer reviewers and editors should improve the completeness of the reporting of blinding procedures, and any test for the success of blinding, for example by endorsing the CONSORT statement.[8] Second, further methodological research is needed to analyse the pros and cons of testing the success of blinding. Such methodological research could enable an international group (e.g. elements of the Cochrane collaboration) to develop methodological guidelines on how best to assess the success of blinding.

Despite the challenges in interpreting results from tests for the success of blinding, our results also point to a need for improving blinding procedures. It is likely that there are substantial practical difficulties in constructing apparently identical treatments in some situations,[52,53] especially in non-pharmacological settings.[54]

In conclusion, we found that tests for the success of blinding are rarely performed. When they are performed, there is considerable variation in the methods of testing and analysis, and a high proportion of incompletely reported and unreported tests. There is an urgent need for improving the methods of evaluating the success of blinding.

KEY MESSAGES

- Blinding can reduce bias in randomized clinical trials, but blinding procedures may be unsuccessful.
- The success of blinding is rarely tested e.g. by asking key trial participants to guess patients' treatments.
- The test methods vary; and the reporting of tests, and test results, is incomplete.
- Seven of 31 trials (23%) reported tests indicating loss of blinding, but in six of the seven cases the risk of a biased trial result was disregarded.
- There is considerable methodological uncertainty about how to assess blinding, and an urgent need for improved methodology and improved reporting.

References

[1] Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;**323:**42–6.

[2] Spanos NP, Burgess CA, Cross PA, MacLeod G. Hypnosis, reporting bias, and suggested negative hallucinations. *J Abnorm Psychol* 1992;**101:**192–9.

[3] Guyatt GH, Pugsley SO, Sullivan MJ *et al*. Effect of encouragement on walking test performance. *Thorax* 1984;**39:**818–22.

[4] Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;**44:**16–20.

[5] Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282:**1054–60.

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

[7] Haahr M, Hróbjartsson A. Who is blinded in randomised clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006;**3:**360–5.

[8] www.consort-statement.org.

[9] Ney PG. Double blind: double talk or are there ways to do better research. *Med Hypothesis* 1986;**21:**119–26.

[10] Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;**328:**432.

[11] Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol* 2005;**58:**1220–6.

[12] Als-Nielsen B, Chen W, Gluud LL, Siersma V, Hilden J, Gluud C. Are trial size and reported methodological quality associated with treatment effects? Observational study of 523 randomised trials. Poster P003. The XII Cochrane Colloqium, Ottawa 2004 (http://cochrane.mcmaster.ca/ottcolloquium.asp).

[13] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;**290:**921–8.

[14] Acworth JP, Purdie D, Clark RC. Intravenous ketamine plus midazolam is superior to intranasal midazolam for emergency paediatric procedural sedation. *Emerg Med J* 2001;**18:**39–45.

[15] Bailey DM, Davies B. Acute mountain sickness; prophylactic benefits of antioxidant vitamin supplementation at high altitude. *High Alt Med Biol* 2001;**2:**21–9.

[16] Belongia EA, Berg R, Liu K. A randomized trial of zinc nasal spray for the treatment of upper respiratory illness in adults. *Am J Med* 2001;**111:**103–8.

[17] Belsito KM, Law PA, Kirk KS, Landa RJ, Zimmerman AW. Lamotrigine therapy for autistic disorder: a randomized, double-blind, placebo-controlled trial. *J Autism Dev Disord* 2001; **31:**175–81.

[18] Brauer LH, Cramblett MJ, Paxton DA, Rose JE. Haloperidol reduces smoking of both nicotine-containing and denicotinized cigarettes. *Psychopharmacology (Berl)* 2001;**159:**31–7.

[19] Brygge T, Heinig JH, Collins P *et al*. Reflexology and bronchial asthma. *Respir Med* 2001;**95:**173–9.

[20] Burke WJ, McArthur-Miller DA. Exploring treatment alternatives: weekly dosing of fluoxetine for the continuation phase of major depressive disorder. *J Clin Psychiatry* 2001;**62 (Suppl 22):**38–42.

[21] Chrusch MJ, Chilibeck PD, Chad KE, Davison KS, Burke DG. Creatine supplementation combined with resistance training in older men. *Med Sci Sports Exerc* 2001;**33:**2111–17.

[22] Curran HV, Kleckham J, Bearn J, Strang J, Wanigaratne S. Effects of methadone on cognition, mood and craving in detoxifying

opiate addicts: a dose-response study. *Psychopharmacology (Berl)* 2001;**154:**153–60.

[23] de Craen AJ, Lampe-Schoenmaeckers AJ, Kraal JW, Tijssen JG, Kleijnen J. Impact of experimentally-induced expectancy on the analgesic efficacy of tramadol in chronic pain patients: a 2×2 factorial, randomized, placebo-controlled, double-blind trial. *J Pain Symptom Manage* 2001;**21:**210–17.

[24] Fukatsu A, Komatsu Y, Senoh H *et al*. Clinical benefits and tolerability of increased fill volumes in Japanese peritoneal dialysis patients. *Perit Dial Int* 2001;**21:**455–61.

[25] Gardner CD, Chatterjee LM, Carlson JJ. The effect of a garlic preparation on plasma lipid levels in moderately hypercholesterolemic adults. *Atherosclerosis* 2001;**154:**213–20.

[26] Gwiazda J, Hyman L, Hussein M *et al*. A randomized clinical trial of progressive addition lenses versus single vision lenses on the progression of myopia in children. *Invest Ophthalmol Vis Sci* 2003; **44:**1492–500.

[27] Hellard ME, Sinclair MI, Forbes AB, Fairley CK. A randomized, blinded, controlled trial investigating the gastrointestinal health effects of drinking water quality. *Environ Health Perspect* 2001; **109:**773–8.

[28] Hutchison KE, Swift R, Rohsenow DJ, Monti PM, Davidson D, Almeida A. Olanzapine reduces urge to drink after drinking cues and a priming dose of alcohol. *Psychopharmacology (Berl)* 2001;**155:**27–34.

[29] Kagan A, Black SE, Duchan JF, Simmons-Mackie N, Square P. Training volunteers as conversation partners using ''supported conversation for adults with aphasia'' (SCA): a controlled trial. *J Speech Hear Disord* 2001;**44:**624–38.

[30] Keaney F, Strang J, Gossop M, Marshall EJ, Farrell M, Welch S, Hahn B, Gonzalez A. A double-blind randomized placebo-controlled trial of lofexidine in alcohol withdrawal: lofexidine is not a useful adjunct to chlordiazepoxide. *Alcohol Alcohol* 2001;**36:** 426–30.

[31] Killen JD, Ammerman S, Rojas N, Varady J, Haydel F, Robinson TN. Do adolescent smokers experience withdrawal effects when deprived of nicotine? *Exp Clin Psychopharmacol* 2001;**9:**176–82.

[32] Laan E, van Lunsen RH, Everaerd W. The effects of tibolone on vaginal blood flow, sexual desire and arousability in postmenopausal women. *Climacteric* 2001;**4:**28–41.

[33] Lintula H, Kokki H, Vanamo K. Single-blind randomized clinical trial of laparoscopic versus open appendicectomy in children. *Br J Surg* 2001;**88:**510–4.

[34] McTavish SF, McPherson MH, Harmer CJ, Clark L, Sharp T, Goodwin GM, Cowen PJ. Antidopaminergic effects of dietary tyrosine depletion in healthy subjects and patients with manic illness. *Br J Psychiatry* 2001;**179:**356–60.

[35] Mendell JR, Barohn RJ, Freimer ML *et al*. Working Group on Peripheral Neuropathy. Randomized controlled trial of IVIg in untreated chronic inflammatory demyelinating polyradiculoneuropathy. *Neurology* 2001;**56:**445–9.

[36] Norheim AJ, Pedersen EJ, Fonnebo V, Berge L. Acupressure treatment of morning sickness in pregnancy. A randomised, double-blind, placebo-controlled study. *Scand J Prim Health Care* 2001;**19:**43–7.

[37] Sackeim HA, Haskett RF, Mulsant BH *et al*. Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA* 2001;**285:**1299–307.

[38] Soares CN, Almeida OP, Joffe H, Cohen LS. Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: a double-blind, randomized, placebo-controlled trial. *Arch Gen Psychiatry* 2001;**58:**529–34.

[39] Teitelbaum JE, Bird B, Greenfield RM, Weiss A, Muenz L, Gould L. Effective treatment of chronic fatigue syndrome and fibromyalgia – a randomized, double-blind, placebo-controlled, intent-to-treat study. *J Chronic Fatigue Syndr* 2001;**8:**3–28.

[40] Turner RB. Ineffectiveness of intranasal zinc gluconate for prevention of experimental rhinovirus colds. *Clin Infect Dis* 2001;**33:**1865–70.

[41] van Wieringen S, Jansen T, Smits MG, Nagtegaal JE, Coenen AML. Melatonin for chronic whiplash syndrome with delayed melatonin onset. *Clin Drug Invest* 2001;**21:**813–20.

[42] Walach H, Rilling C, Engelke U. Efficacy of Bach-flower remedies in test anxiety: a double-blind, placebo-controlled, randomized trial with partial crossover. *J Anxiety Disord* 2001;**15:**359–66.

[43] Wisner KL, Perel JM, Peindl KS, Hanusa BH, Findling RL, Rapport D. Prevention of recurrent postpartum depression: a randomized clinical trial. *J Clin Psychiatry* 2001;**62:**82–6.

[44] Wood A, Trainor G, Rothwell J, Moore A, Harrington R. Randomized trial of group therapy for repeated deliberate self-harm in adolescents. *J Am Acad Child Adolesc Psychiatry* 2001;**40:**1246–53.

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

[46] Sackett D. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;**328:**1136.

[47] James KE, Lee KK, Kraemer HC, Fuller RK. An index for assessing blindness in a multicenter clinical trial: disulfiram for alcohol cessation-a VA cooperative study. *Stat Med* 1990;**15:**1421–34.

[48] Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Control Clin Trials* 2004;**25:**143–56.

[49] Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to publications. *JAMA* 2004;**291:**2457–65.

[50] Rees JR, Wade TJ, Levy DA, Colford JM Jr, Hilton JF. Changes in belifs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Con Clin Trials* 2005;**26:**25–37.

[51] Walter SD, Awashti S, Jeyaseelan L. Pre-trial evaluation of the potential for unblinding in drug trials: a prototype example. *Con Clin Trials* 2005;**26:**459–68.

[52] Hill LE, Nunn AJ, Fox W. Matching quality of agents employed in ''double-blind'' controlled clinical trials. *Lancet* 1976;**1:**352–6.

[53] Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, Ravaud P. Methods of blinding in reports of randomized controlled trials assessing pharmacological treatments. A systematic review. *PLoS Medicine* 2006;**3:**e425.

[54] Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 2004;**57:**543–50.

Appendix Characteristics of the trials testing the success of blinding ($N = 31$)

Trial	Design	Condition	Compared groups	1. Outcome	Explicitly blinded	Tested	Author's conclusion
Acworth	Parallel $n = 53$ 'Single blind'	Sedation	Kettamine iv Midazolam iv Midazolam iv	Mean sedation score (0–5)	Data coll	Data coll	Unblinding 'not supported by the frequency with which scorers' guessed the type of sedation
Bailey	Parallel $n = 18$ 'Double blind'	Mountain sickness	Antioxidant caps Placebo caps	Mean Lake Louise AMS score	Patients	Patients	'. . . , thus confirming the effectiveness of our blinding protocol'
Belongia	Parallel $n =$ ns 'Double blind'	Upper respiratory illness	Zink spray Placebo spray	Number of days with symptoms	Patients All study personel	Patients	No explicit conclusion, but $P = 0.59$ (Chi^2)
Belsito	Parallel $n = 37$ 'Single blind'	Autism	Lamotrigene tabl Placebo tabl	Mean autism behaviour checklist (AUBC) score	HCP Data coll Data analysts	HCP Data coll	'Outcome assessors [data coll] . . . were unable to predict who was in which group . . .'
Brauer	Crossover $n =$ ns 'Double blind'	Smoking	Haloperidol 1 mg Haloperidol 2 mg Placebo caps	Mean number of cigarettes smoked	Patients	Patients	ns
Brygge	Parallel $n = 40$ 'Double blind'	Asthma	Reflexology massage Placebo massage	Mean symptom score (0–4)	Patients Data coll	Patients	'. . . a clear tendency toward correctly guessing the treatment received was apparent . . .'
Burke	Parallel $n = 70$ 'Double blind'	Depression	Fluoxetine 20 mg/d Fluoxetine 60 mg/w Placebo caps	Mean Montgomery-Asberg depression rating scale	Patients	Patients	'Fischer exact tests of patients' ability to correctly identify treatment assignment at study end revealed no significant difference . . .'
Chrusch	Parallel $n = 33$ 'Double blind'	Muscle atrophy	Creatine mixt Placebo mixt	Mean leg press in kg	Patients	Patients	'No significant difference was noted in treatment identification . . .'
Curran	Parallel $n = 24$ 'Double blind'	Opiate addiction	Methadone syrup Placebo syrup	Mean Rivermead behavioural memory test (prose recall scores)	Patients	Patients	'Therefore, patients could not differentiate between methadone and placebo treatments'
de Craen	Parallel (2×2) $n = 112$ 'Double blind'	Chronic pain	Tremadol caps Placebo caps	Mean pain decrease (10 cm VAS)	Patients All trial personnel	Patients	'In our trial, significantly more patients in the placebo group correctly guessed their treatment allocation . . .'
Fukatsu	Crossover $n = 22$ 'Double blind'	Peritoneal dialysis	Fill volume 2.5 L Fill volume 2.0 L Fill volume 1.5 L	Mean discomfort scores	Patients Some staff members	Patients	ns
Gardner	Parallel $n = 53$ 'Double blind'	Hypercholesterolemia	Garlic 0.5 g/d Garlic 1 g/d Placebo tabl	Mean concentration (mg/dl) of plasma LDL-cholesterol	Patients Study staff Laboratory personel Investigators	Patients	'The blinding technique proved to be effective'

(continued)

Appendix Continued

Trial	Design	Condition	Compared groups	1. Outcome	Explicitly blinded	Tested	Author's conclusion
Gwiazda	Parallel $n = 469$ 'Double blind'	Myopia progression	Progressive add lenses Single vision lenses	Mean change in spherical equivalent refractive error	Patients Parents Some trial staff	Patients	ns
Hellard	Parallel $n = 600$ 'Double blind'	Gastro enteritis	Water treatment units Placebo units	Proportion of cases with gastroenteritis (based on questionnaire)	Patients Investigators Plumbers	Patients	'The study was successfully blinded . . .'
Hutchison	Parallel $n = 26$ 'Patient blind'	Social drinkers	Olanzapine caps Placebo caps	Mean alcohol urge (11 point) score	Patients	Patients	ns
Kagan	Parallel $n = 40$ 'Single blind'	Volunteers for aphasic patients	Supported conversation No treatment	Mean score on Measure of Supported Conversation for Adults with Aphasia	Data coll	Data coll	'. . . the rater [data coll] was usually correct . . . in guessing about the status of the rated interview'
Keaney	Parallel $n = 80$ 'Double blind'	Alcohol withdrawal	Lofexidine tabl Placebo tabl	Mean alcohol withdrawal scale (AWS)	Patients Data analysts Research staff Clinical staff	Patients	'Patient's self-report with respect to which medication they had taken revealed no significant difference between the two groups'
Killen	Parallel $n = 105$ 'Double blind'	Nicotine withdrawal	Nicotine patch Placebo patch	Mean modified Fagerstrom tolerance questionnaire scores	Patients Staff	Patients	No explicit conclusion, but $P = 0.18$ (Chi^2)
Lintula	Parallel $n = 73$ 'Single blind'	Appendicitis	Laparoscopic op Open op	Mean pain score	Patients Parents Research nurses Pediatric surgeons	Patients HCP	ns
Laan	Crossover $n = 44$ 'Double blind'	Sexual dysfunction	Tibolone Placebo	Mean weekly frequency of sexual araousability	Patients	Patients	ns
McTavish	Parallel $n = 20$ 'Blind'	Mania	Tyrosine-free mixt Tyrosine mixt	Mean % decrease from baseline on Beigel manic state rating scale	Clinical rater	Data coll ('rater')	'. . . whereby prediction of the mixture administered by the blinded rater was correct for 75% of patients, $P < 0.05$'
Mendell	Parallel $n = 53$ 'Double blind'	CIDP	Ig iv Placebo iv	Mean change in average muscle score	Patients	Patients	'. . . strongly supporting that the blind had been maintained'

Nordheim	Parallel $n = 97$ 'Double blind'	Morning sickness	P6 acupressure Placebo acupressure	Proportion of women not improving (on overall symptom score, 1–5)	Patients Investigators	Patients	'... participants in the control group guessed better what type of band they had used'
Sackeim	Parallel $n = 84$ 'Double blind'	Depression	Nortriptyline + Li caps Nortriptyline caps Placebo caps	Proportion of patients with relapse (based on Hamilton Rating Scale for Depression)	Patients Treatment team Outcome assessor Data analysts	Patients	'... patient blinding was imperfect ...'
Soares	Parallel $n = 50$ 'Double blind'	Depression	Estradiol patch Placebo patch	Mean MADRS depression scale ratings	Patients Psychiatric rater	Patients Data coll	'... the assessment of blindness ... suggested that it was acceptable'
Teitel-Baum	Parallel $n = 72$ 'Double blind'	Chronic fatigue	Several 'active' drugs Placebos	Mean overall response score	Patients Treating physician	Patients	'... suggests that unblinding did not occur'
Turner	Parallel $n = 105$ 'Double blind'	Common cold	Zink spray Placebo spray	Proportion with rhinovirus infection	Patients	Patients	No explicit conclusion, but $P = 0.29$ (Fischer)
Walach	Parallel $n = 61$ 'Double blind'	Test anxiety	Bach-flower solution Placebo solutions	Mean reduction in test anxiety inventory (TAI-G) score	Patients Investigators	Patients	'Thus, the blindness ... remained unchallenged to the end'
Wieringen	Parallel $n = 81$ 'Double blind'	Chronic whiplash	Melatonin tabl Placebo tabl	Mean lights-off time (Sleep onset diary)	Patients Investigators	Patients	ns
Wisner	Parallel $n = 56$ 'Double blind'	Postpartum depression	Nortriptyline caps Placebo caps	Proportion of patients with recurrent depression (based on Hamilton Rating Scale for Depression)	Patients Investigator Study coordinator Symptom monitor Side effect monitor	Patients Data coll ('Side effect monitor') Investig	'The nurse who evaluated side effects was able to guess the drug assignment ... better than chance ...'
Wood	Parallel $n = 63$ 'Double blind'	Deliberate self-harm	Group psychotherapy Routine care	Proportion of cases with repeated self-harm		Data coll	'... responses were no better than chance'

Note: n: number of included patients; iv: intra venous; in: intra nasal; caps: capsules; ns: not stated; tabl: tablets; HCP: health care providers; d: day; w: week; mixt: mixture; add: addition; op: operation; Data coll: data collectors; Ig: immuno globulin; Investig: primary investigator and study coordinator.