Papers

Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials

Dean Fergusson, Kathleen Cranley Glass, Duff Waring, Stan Shapiro

Abstract

Objective To examine the reporting and success of double blinding in a sample of randomised, placebo controlled trials from leading general medicine and psychiatry journals.

Methods Identification of placebo controlled, randomised controlled trials from prespecified general medical and psychiatric journals indexed on Medline between 1 January 1998 and 1 October 2001, from which a random sample of 200 randomised clinical trials was chosen, of which 191 trials were evaluated.

Results Only seven of the 97 (7%) general medicine trials provided evidence on the success of blinding, with five reporting that the success of blinding was imperfect. In trials from psychiatric journals, the success of blinding was reported in eight of the 94 trials, with four reporting that the blinding was imperfect. Overall, only four of the 191 (2%) trials assessed blinding in the participants and either the outcome assessors or the investigators.

Conclusions The current lack of reporting on the success of blinding provides little evidence that success of blinding is maintained in placebo controlled trials. Trialists and editors should make a concerted effort to incorporate, report, and publish such information and its potential effect on study results.

Introduction

Although the definition of double blind varies,[1] we consider a trial to be double blind when the patient, investigators, and outcome assessors are unaware of the patient's assigned treatment throughout the conduct of the trial.[2] Placebos are commonly used as an inactive treatment to achieve double blinding. Active placebos, with which symptoms or side effects are imitated, can also be used. Placebos are justly used when no existing effective treatment is available. Sometimes, placebos are proposed instead of a standard existing treatment or standard care to ensure assay sensitivity. That is, to demonstrate the effectiveness of a new treatment, it must be demonstrated against a "clean" control. The argument is that although the new treatment may be found to be as effective or more effective than standard treatment in a clinical trial, both treatments may very well be ineffective. Assay sensitivity is the ability of a trial to distinguish effective interventions from ineffective interventions. It depends on the effect size that is to be detected. As such, the investigators need to know the anticipated effects of the control intervention. It is argued that placebos are the ideal choice as their anticipated benefits are known to be marginal. This argument is predicated on the belief that participants, investigators, and outcome assessors remain blinded to the treatment assignment. If the blinding of the placebo arm is not effective then the protection against expectation effects, biased assessment, contamination, and co-intervention are all lost. The observed superiority of a new treatment over placebo could merely be a consequence of loss of this control—and an ineffective new treatment would spuriously seem to be superior. Because of the importance of the success of blinding, the Consolidated Standards for Reporting of Trials (CONSORT) Group has explicitly incorporated the issue. Section 11(b) of the CONSORT statement states that the success of blinding is to be reported in the publication.[3]

It is not sufficient that trials describe themselves as double blind. It is also important that the efficacy of the blinding is actually assessed. In other words, an assessment of the face validity of the double blinding is needed. To assess the reporting and success of double blinding, we chose a random sample of randomised, placebo controlled trials from leading journals in general medicine and psychiatry. Although we have focused on placebo controlled trials, the issues discussed also arise in double blind trials with active controls.

Methods

For this study we selected five of the top general medical journals (*JAMA*, *New England Journal of Medicine*, *BMJ*, *Lancet*, and *Annals of Internal Medicine*) and four of the top journals in psychiatry (*Archives of General Psychiatry*, *Journal of Clinical Psychiatry*, *British Journal of Psychiatry*, and *American Journal of Psychiatry*). Our Medline search used publication type "randomised controlled trial" and the MeSH term "placebo-controlled" to identify placebo controlled randomised trials that were indexed on Medline between the dates of 1 January 1998 and 1 October 2001 and published in these nine journals. All citations from this search were then consecutively numbered, and a random number generator was used to select 100 trials from the general medicine journals and 100 trials from the psychiatry journals. We reasoned that 100 trials from each discipline was a manageable number to abstract and an adequate number to obtain a good estimate of the number of trials reporting the success of blinding, and we performed no formal sample size calculations.

Data abstraction forms were developed and included document identification, the type of interventions, type of placebo, matching characteristics of placebo to intervention, who was blinded, and the evidence of successful blinding. A trial indicating that a "similar" placebo was used but did not specify how it was similar was scored as "not mentioned" (our rationale is that the term "similar" is vague and therefore inadequate). The page number and location for each data item was also recorded. Six people abstracted all the data. At least two people independently

Table 1 Type of intervention and placebo. Values are number of trials

	General medicine	Psychiatry
Type of intervention*		
Pharmacological	83	89
Surgical	0	0
Behavioural	1	2
Device	0	3
Physiotherapeutic	3	0
Nutritional supplement	9	0
Other	4	3
Total	100	97
Type of placebo†		
Not reported	16	38
Capsule	18	28
Tablet	20	16
Injection	23	5
Sham device	2	1
Sham procedure	2	0
Oral solution	9	3
Inhaler	2	0
Other	9	3
Total	101	94

*3 trials in general medicine and 3 trials in psychiatry had more than one intervention.
†4 trials in general medicine and 1 trial in psychiatry had more than one type of placebo.

abstracted data for each study. Either consensus or a third party resolved any differences. All data were entered into an electronic database (Microsoft Excel 2000).

Results

The Medline search identified a total of 473 randomised controlled trials in the general medicine literature and 192 trials in the psychiatric literature, from which we randomly chose 100 trials from general medicine and 100 trials from psychiatry. Nine trials were removed from further analysis as they were not placebo controlled trials despite being identified as such in the systematic literature search. Thus, we evaluated 97 trials from the general medicine literature and 94 trials from psychiatry literature.

General medicine
Table 1 provides information on the type of interventions and placebos used in the 97 trials in general medicine. Eighty three per cent of all interventions were pharmacological. Nutritional supplements (9%) were the second most frequent intervention used. Sixteen of the 97 trials did not report the type of placebo used. Of trials that reported the type of placebo, an injection (either subcutaneous or intramuscular) was the most common (23; 27%), followed by tablet (20; 24%) and capsule (18; 21%).

The matching characteristics of placebo to the intervention was reported in 51 (53%) of the trials; one trial (1%) also reported the dissimilarity between placebo and intervention (table 2). Appearance was the characteristic most often reported by investigators (46 of 51 trials), followed by taste (9 of 51 trials).

Only seven of the 97 trials (7%) provided evidence on the success of blinding (table 3). [4–10] All seven trials assessed the success of blinding in study participants. One trial assessed the success of blinding in individuals assessing study outcome.[6] All seven trials presented a method for assessing blinding. Five of the trials presented blinding data for each trial arm, one trial presented overall aggregated data only, and one trial provided no data. Five trials reported that the success of blinding was imperfect.[4 6–8 10] The trial that did not present blinding data described blinding as successful without further comment.[9] The

Table 2 Number of studies reporting matching of characteristics of placebo to intervention

Matching characteristic	General medicine (n=97)	Psychiatry (n=94)
Not reported	46	64
Reported:	51	30
Appearance	46	27
Taste	9	3
Smell	2	3
Side effects	0	0
Feel or touch	4	2
Other	4	3
Trials with >1 characteristic reported	12	7
Discordance reported	1	2

trial that reported aggregated blinding data did not comment, qualitatively, or provide statistical tests of success of blinding.[5]

Psychiatry
Most psychiatry trials used pharmacological interventions (table 1). Over 40% of the trials did not report the type of placebo used. Of the placebos reported, 78% were either a capsule or tablet.

The matching characteristics between intervention and placebo were reported in 30 (32%) of trials, with appearance being the most often reported (table 2). Eight of the 94 trials reported evidence on successful blinding (table 4).[11–18] Of these, six assessed the success of blinding in patients.[12 14–18] Two studies provided blinding data for both subjects and outcome assessors[16 17]; one study reported blinding data for both treatment administrators and outcome assessors,[11] and one study provided data for treatment administrators only.[13] Six of the eight studies presented a method for blinding assessment in the Methods or Results section of the article and the other two presented it in the Discussion section. Four of the trials presented blinding data broken down by treatment allocation[12 14 15 17]; one trial presented aggregated data and did not provide data broken down by treatment allocation[13]; and two presented no data on blinding.[11 16] Of the eight trials, the blinding was reported as less than optimal in four.[11 14 15 18]

Discussion

The quality of reporting in clinical trials has evolved. Over the years, trialists have been held more accountable and responsible for the quality of trial reporting. This evolution began with the need for reporting the numbers of patients screened, enrolled, randomised, and analysed,[19] and progressed to the reporting of patient withdrawals and its importance for the analysis and interpretation of study results.[20] Building on this progress, there is a need for trialists and journals routinely to report the methods of blinding and the subsequent success of this blinding.[21]

Our examination of the success of blinding challenges the notion that placebo controlled trials inherently possess assay sensitivity. Clearly, there is a failure among investigators and journals in reporting the success of blinding. Only 15 of the 191 trials (8%) provided such information, be it qualitative or quantitative. Of the 15 trials, only five trials reported that blinding was successful,[9 12 13 16 17] and of these, three did not present any quantitative data analysis to support their claim.[9 13 16]

Only four trials assessed blinding in both the participants and either the outcome assessors or the investigators.[6 12 16 17] Thus, the face validity of the double blinding was only reported in four of the 191 articles (2%). This deficiency in reporting

Table 3 Reporting of blinding assessment in 97 trials in general medicine

Study	Outcome and patient population	Intervention	Individuals guessing (proportion guessing)	% correct guesses			% undecided			Authors' statements on success of blinding
				Active	Placebo	Total	Active	Placebo	Total	
Sackeim et al[4]	Relapse in major depression	Nortriptyline; nortriptyline plus lithium	Subjects (63/ 73*)	65; 24†	50	NK	NK	NK	NK	a) "Analysis yielded a modest association between treatment assignment and the patient's guesses (χ^2=9.68, P=0.05)" b) "While the patient blinding was imperfect, relapse status was a more powerful determinant of the guesses"
Rowe et al[5]	Global wellness in chronic fatigue syndrome	Fludrocortisone	Subjects (NK)	NK	NK	42	NK	NK	NK	Not stated
Apfel et al[6]	Neuropathy in people with type 1 and 2 diabetes with polyneuropathy	Recombinant human nerve growth factor	Subjects (NK)	33	49	NK	NK	NK	NK	"Although the rates of correct identification was less than 50%, the availability of an "unknown" category resulted in a statistically significant association between the actual treatment received and the opinion about the treatment received (P<0.05)"
			Outcome assessors (NK)	34	42	NK	NK	NK	NK	
Von Schacky et al[7]	Artherosclerosis in patients with diagnosed coronary artherosclerosis	Fish oil concentrate	Subjects (175/223)	24	11	18	70	78	74	"χ^2 test; P=0.06" (no interpretation provided)
Sandler et al[8]	Gastrointestinal symptoms in consumers	Olestra chips	Subjects (3055/3250)	39	12	NK	NK	NK	58	"An interesting finding was the association of gastrointestinal symptoms with the type of chips that participants thought they were eating. Participants who thought they were eating olestra chips reported gastrointestinal symptoms approximately 50% more often than participants who believed that they were eating regular chips."
Blondel et al[9]	Abstinence in smokers	Nicotine nasal spray	Subjects (NK)	NK	NK	NK	NK	NK	NK	"Blinding among participants was successful. At the 1 year follow up we found no significant relation between type of treatment and the participant's responses, which proved they had been unable to guess their treatment"
Shlay et al[10]	Pain scores in HIV associated, symptomatic, lower extremity peripheral neuropathy	Amitryptiline	Subjects (111/136)	67	58	NK	16	9	NK	a) "P<.001" b) "The indication that the blinding was not maintained also confirms the lack of efficacy because unblinding tends to bias toward a hypothesized intervention."

NK=not known.

*73 of the 84 participants randomised to the placebo, nortriptyline, or nortriptyline plus lithium arm completed the study.

†First value is for nortriptyline arm and the second value is for nortriptyline plus lithium arm.

translates into a paucity of evidence that a placebo ensures a "clean" control. Furthermore, the quality of evidence in the few studies that reported on the success of blinding is weak on two fronts: the quality of the data and the evidence that blinding was successful. The success of blinding was described as less than optimal in nine of the 14 trials that reported on blinding, and of the five trials that reported that blinding was maintained, only two provided data to support their claim.[12 17] Unfortunately, when we examined the data and analysis provided by these two trials we found that their claim of success is debatable.

We would like to see Item 11b of CONSORT revised to require the assessment of blinding for all double blind randomised trials. Trialists have an ethical responsibility to justify the use of a placebo for blinding purposes in their research protocol and informed consent procedures. Thus, it seems reasonable to suggest that an assessment of the success of blinding is necessary. If blinding is not assessed, we may delude ourselves as to exactly what information we gain from incorporating a placebo comparison. Although all trials should assess blinding, the types of trials that will particularly benefit are trials with subjective outcomes or outcomes reported by patients

(for example, quality of life instruments), or trials where the side effects are well known. Even though there may be problems with analysing and interpreting the results of success, this does not provide a rationale for not doing it. Clearly, the lack of successful blinding can bias observed estimates of effect. Although this bias is differential, its direction may not be easily ascertained. We might anticipate that evidence of unsuccessful blinding in a "double blind" active versus placebo trial would result in a positive bias and hence lead to an overestimate of the treatment effect. However, unblinded patients receiving placebo may seek other treatments, especially if there is established effective treatment available, and this makes the extent and even the direction of bias difficult to determine.

We believe that trialists need to report a minimum set of information. This includes the counts of all patients allocated to each treatment; the counts of patients who guess treatment assignment by the group to which they were allocated; the counts of correct guesses and those who are undecided; the analytical methods and results used to assess success of blinding; and the author's interpretation of the efficacy of blinding and the effect on study results. The data abstracted for this study show a

Table 4 Reporting of blinding assessment in 94 trials in psychiatry

Study	Outcome and patient population	Intervention	Individuals guessing (proportion guessing)	% correct guesses			% undecided			Author's statements on success of blinding
				Active	Placebo	Total	Active	Placebo	Total	
Wisner et al[11]	Depression in patients with major postpartum depression	Nortryptiline	Outcome assessors (NK)	NK	NK	NK	NK	NK	NK	"Nurse who evaluated side effects was able to guess better than chance (κ=0.47, P=0.01)"
			Other personnel (NK)	NK	NK	NK	NK	NK	NK	"κ from −0.02 to 0.18" (no interpretation provided)
Warner et al[12]	Bereaved individuals	Diazepam	Subjects (22/30)	75	64	NK	NK	NK	NK	"Subjects were not aware of their allocation (Fishers exact test P=0.18)"
			Interviewers (NK)	NK	NK	NK	NK	NK	NK	"There was no evidence that the interviewers were aware of the treatment allocation"
Ben Zion et al[13]	Anxiety in healthy volunteers	Metergoline	Physician (3/3)	NK	NK	55	NK	NK	NK	"Blinding maintained"
Himle et al[14]	Anxiety in people with social phobia	Alcohol	Subjects (40/40)	80	50	65	NK	NK	NK	"χ^2=2.75, df=1, P=0.10, N=40"; "The findings of the present study related to the effect believing that one received alcohol are complex and difficult to interpret, but finding significant differences on two different measures of anxious response suggests that these expectancies did have an effect"
Stoll et al[15]	Remission in bipolar disorder	ω3 fatty acids	Subjects (NK)	86	63	NK	NK	NK	NK	"Although in some cases the guess was based on a fishy aftertaste, in many cases it was based on the patient's perceived clinical response (or lack thereof in the placebo group)"
Heresco-Levy et al[16]	Symptom scores in patients with schizophrenia	Glycine	Outcome assessors (NK)	NK	NK	NK	NK	NK	NK	"The raters, patients, and their families were unaware of and could not determine the study drug assignment by taste or otherwise."
			Subject's family (NK)	NK	NK	NK	NK	NK	NK	
			Subjects (NK)	NK	NK	NK	NK	NK	NK	
Schneier et al[17]	Symptom or function scores in people with social phobia	Moclobemide	Subjects (NK)	62	45	NK	NK	NK	NK	"χ^2=0.19, df=1, P=0.90" (no interpretation provided)
			Outcome assessors (NK)	25	41	NK	NK	NK	NK	"χ^2=2.06, df=1, P=0.40" (no interpretation provided)
Young et al[18]	Symptoms in patients with premenstrual dysphoric disorder	Sertraline	Subjects (NK)	100*	100*	100*	NK	NK	NK	"Subjects were queried at the end of this study to assess the effectiveness of blinding of the treatment order and all were able to correctly identify which treatment was received in each period."

NK=not known.
*As determined by the author's statement on success of blinding.

What is already known on this topic

Placebo controls are commonly used in randomised trials to blind investigators, outcome assessors, and patients to treatment assignment

Placebo controls have been advocated instead of existing effective treatment because they ensure assay sensitivity

Unsuccessful double blinding results in a differential bias of effect measures

What this study adds

The success of blinding is not well reported

The success of blinding in trials that do report is often poor

Little evidence exists that placebos provide assay sensitivity

substantial lack of reporting with respect to these minimum, essential items, as illustrated by the number of vacant fields in tables 3 and 4.

The current lack of reporting on the success of blinding provides little evidence that success of blinding is maintained in placebo controlled trials. Trialists and editors need to make a concerted effort to incorporate, report, and publish such information and its potential effect on study results. The efficacy of the blinding cannot be assumed on theoretical grounds. We need evidence before we can assert that assay sensitivity exists in randomised, double blind, placebo controlled trials.

1 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-3.
2 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254-9.
3 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
4 Sackeim HA, Haskett RF, Mulsant BH, Thase ME, Mann JJ, Pettinati HM, et al. Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA* 2001;285:1299-307.

5 Rowe PC, Calkins H, DeBusk K, McKenzie R, Anand R, Sharma G, et al. Fludrocortisone acetate to treat neurally mediated hypotension in chronic fatigue syndrome: a randomized controlled trial. *JAMA* 2001;285:52-9.

6 Apfel SC, Schwartz S, Adornato BT, Freeman R, Biton V, Rendell M, et al. Efficacy and safety of recombinant human nerve growth factor in patients with diabetic polyneuropathy: a randomized controlled trial. rhNGF Clinical Investigator Group. *JAMA* 2000;284:2215-21.

7 Von Schacky C, Angerer P, Kothny W, Theisen K, Mudra H. The effect of dietary omega-3 fatty acids on coronary atherosclerosis. A randomised, double-blind, placebo-controlled trial. *Ann Intern Med* 1999;130:554-62.

8 Sandler RS, Zorich NL, Filloon TG, Wiseman HB, Lietz DJ, Brock MH, et al. Gastrointestinal symptoms in 3181 volunteers ingesting snack foods containing olestra or triglycerides. A 6-week randomised, placebo-controlled trial. *Ann Intern Med* 1999;130:253-61.

9 Blondal T, Gudmundsson LJ, Olafsdottir I, Gustavsson G, Westin A. Nicotine nasal spray with nicotine patch for smoking cessation: randomised trial with six year follow up. *BMJ* 1999;318:285-8.

10 Shlay JC, Chaloner K, Max MB, Flaws B, Reichelderfer P, Wentworth D, et al. Acupuncture and amitriptyline for pain due to HIV-related peripheral neuropathy: a randomised controlled trial. Terry Beirn Community Programs for Clinical Research on AIDS. *JAMA* 1998;280:1590-5.

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

12 Warner J, Metcalfe C, King M. Evaluating the use of benzodiazepines following recent bereavement. *Br J Psychiatry* 2001;178:36-41.

13 Ben Zion IZ, Meiri G, Greenberg BD, Murphy DL, Benjamin J. Enhancement of CO_2-induced anxiety in healthy volunteers with the serotonin antagonist metergoline. *Am J Psychiatry* 1999;156:1635-7.

14 Himle JA, Abelson JL, Haghightgou H, Hill EM, Nesse RM, Curtis GC. Effect of alcohol on social phobic anxiety. *Am J Psychiatry* 1999;156:1237-43.

15 Stoll AL, Severus WE, Freeman MP, Rueter S, Zboyan HA, Diamond E, et al. Omega 3 fatty acids in bipolar disorder: a preliminary double-blind, placebo-controlled trial. *Arch Gen Psychiatry* 1999;56:407-12.

16 Heresco-Levy U, Javitt DC, Ermilov M, Mordel C, Silipo G, Lichtenstein M. Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia. *Arch Gen Psychiatry* 1999;56:29-36.

17 Schneier FR, Goetz D, Campeas R, Fallon B, Marshall R, Liebowitz MR. Placebo-controlled trial of moclobemide in social phobia. *Br J Psychiatry* 1998;172:70-7.

18 Young SA, Hurt PH, Benedek DM, Howard RS. Treatment of premenstrual dysphoric disorder with sertraline during the luteal phase: a randomised, double-blind, placebo-controlled crossover trial. *J Clin Psychiatry* 1998;59:76-80.

19 Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;301:1410-2.

20 Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57:6-15.

21 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742-4.

Ottawa Health Research Institute, Clinical Epidemiology Program, 501 Smyth Road, Box 201, Ottawa, ON, Canada K1H 8L6
Dean Fergusson *scientist*

Departments of Human Genetics & Pediatrics and Biomedical Ethics Unit, McGill University, Montreal, QC, Canada
Kathleen Cranley Glass *professor*

Department of Epidemiology and Biostatistics, McGill University
Stan Shapiro *professor*

Research Ethics and Regulation Group, Faculty of Law, University of Toronto, Toronto, ON, Canada
Duff Waring *research associate*

Correspondence to: D Fergusson dafergusson@ohri.ca

Amendment

This is Version 2 of the paper. In this version, the references in tables 3 and 4 have been corrected. They now start with reference 4 and end with reference 18 [in the previous version they ran from reference 2 to reference 16].