Published in final edited form as:

Cephalalgia. 2006 August ; 26(8): 973–982.

Blindness and bias in a trial of antidepressant medication for chronic tension-type headache

KA Holroyd^{1,3}, G Tkachuk^{1,3}, F O'Donnell³, and GE Cordingley²

¹Department of Psychology, Westerville, OH, USA

²College of Osteopathic Medicine, Ohio University, Athens, Westerville, OH, USA

³Headache Treatment & Research, Westerville, OH, USA

Abstract

This study aimed to examine penetration of the blind in a randomized, placebo-controlled trial. Neurologists' ratings of improvement and medication side-effects, participants' ratings of improvement and daily diary recordings of headaches were assessed along with participants' and neurologists' guesses about treatment group placement in participants who completed at least 3 months of treatment ($N = 169$). Despite blinding, treating neurologists successfully identified the medication condition for 82% of participants receiving medication only; trial participants accurately identified their medication condition when receiving active medication (77% of participants), but not when receiving placebo. Concurrent stress-management therapy reduced, but did not eliminate penetration of the blind. Irrespective of drug condition, when participants were improved they were judged to be on active medication and when unimproved they were judged to be on placebo. However, neurologists' ratings of improvement, participants' reports of improvement and daily headache recordings yielded equivalent outcomes. Penetration of the blind needs to be assessed, not assumed in clinical trials in headache. However, penetration of the blind did not produce a prodrug bias as has been asserted by critics. Better methods of assessing and quantifying blindness are needed.

Keywords

Bias; clinical trial; double blind; headache; methodology

Introduction

The double blind methodology in clinical trials aims to reduce bias from the expectations of study staff and participants (1). Penetration of the blind may lead trial participants or treating physicians to expect improvement for participants accurately perceived to be on active medication and to expect a lack of improvement for participants accurately perceived to be on placebo. Expectations then may influence actual improvement, participant reports of improvement, or assessors' ratings, thereby biasing trial results (2–5).

Blindness may be penetrated in at least two ways. The presence or absence of side-effects may provide clues as to a participant's treatment condition. In fact, detailed specification of expected side-effects in contemporary informed consent materials may increase the likelihood participants are able to use side-effects as a cue to penetrate the blind (6). If side-effects lead participants and treating physicians or evaluators to penetrate the blind early in treatment, the trial may effectively be unblinded (7–12). Trial participants and clinicians may also correctly

identify study medication as an active drug if improvement occurs and as placebo if there is no improvement (8,13). If improvement leads participants or clinicians to penetrate the blind, the treatment effect has occurred prior to penetration of the blind and bias is unlikely (14).

Few studies report tests of whether double-blind procedures actually maintain blindness (15). Where the integrity of the blind has been assessed, however, trial participants and treating physicians or evaluators are often found to identify participants' treatment condition correctly at rates greater than chance (7,8,10,13,16–23). These observations have led some investigators to assert that trials in psychiatry, where these findings have been obtained, have a prodrug bias (23–29). In particular, it is asserted that evaluations by unblinded prodrug clinicians are biased, while no placebo response occurs in unblinded trial participants receiving placebo, so estimates of drug effects are inflated. However, only limited information is available about the causes and consequences of this penetration of the blind. In particular, it has yet to be demonstrated that trials where the blind has been penetrated necessarily have a prodrug bias (14,18,30).

To our knowledge, no trial evaluating treatments for headache has examined the integrity of the double blind. We examined penetration of the blind by both study neurologists and trial participants in a double-blind trial of tricyclic antidepressant medication and cognitive-behavioural stress-management therapy for chronic tension-type headache (31). We also examined factors that might lead to penetration of the blind, in particular, relationships between both headache improvement and side-effects and penetration of the blind. In addition, we attempted to determine if penetration of the blind biased treating neurologists' and trial participants' ratings of improvement. Finally, we asked if penetration of the blind was less likely when trial participants received concurrent stress management therapy for headaches than when trial participants received drug therapy alone. We reasoned that where either stress management or drug treatment might produce improvements in headaches and altered bodily sensations, these cues would provide an unreliable indication of a trial participants' drug condition.

Methods

Treatment of chronic tension-type headache trial design

The Ohio University Institutional Review Board granted approval for this trial. All participants provided written informed consent. The trial compared four treatments for chronic tension-type headache in a randomized ($N = 203$) parallel group design: tricyclic antidepressant medication (AM), placebo (PL), placebo plus stress-management therapy (PL + SMT) and antidepressant medication plus stress-management therapy (AM + SMT). The trial included a 1-month baseline assessment phase, a 2-month treatment/dose adjustment phase, in which the antidepressant medication dose was adjusted and stress-management therapy was administered, and 3- and 8-month evaluations. Data reported here were collected through the 3-month evaluation where blindness was assessed.

Participants

Inclusion criteria were age between 18 and 65 years and an International Headache Society (IHS) (32) diagnosis of chronic tension-type headache at separate evaluations by a neurologist and by a second evaluator. Exclusion criteria included: IHS diagnosis of analgesic abuse headache (32), current use of prophylactic medication for tension-type headache; migraine headache more than 1 day per month; and pain problem (e.g. arthritis) other than headache as the primary pain problem.

Of the 203 participants who were randomized, 169 (AM = 48; PL = 38; AM + SMT = 45; PL + SMT = 38) completed the 3-month evaluation where blindness was assessed (see trial report

for attrition details; (31)). These 169 participants were primarily female (75%), white (96%), on average 38 years old, with tension headaches occurring, on average, 26 days per month with mean disease duration of 13 years.

Treatments

Medication/placebo—The double-blind AM protocol attempted to maximize the efficacy and tolerability of AM by beginning medication one (amitriptyline hydrochloride, or matched placebo) at a low dose (12.5 mg amitriptyline), providing a second medication option (nortriptyline hydrochloride, or matched nortriptyline placebo) for participants who failed to tolerate medication one, dosing to the recommended dose levels for the treatment of chronic tension-type headache (75–100 mg amitriptyline, or equivalent matched placebo, or 50–75 mg nortriptyline, or equivalent matched nortriptyline placebo) or to tolerance, and incorporating an adherence intervention (33–35). Details of the protocol including the dosing schedule, final dose levels and adherence information can be found in the original trial report (31).

Stress management—In SMT, relaxation and cognitive coping skills were also introduced in separate counselling sessions at the three medication adjustment clinic visits; homework and participants' application of relaxation and cognitive coping skills in the management of headaches were guided by a manual with eight accompanying audiotapes (36). Details of this treatment protocol can be found in the original trial report (31) and a treatment manual can be found in Lipchik and colleagues (37).

Measures

Headaches—Participants recorded headaches in a daily diary (31). Headache activity was recorded four times a day using an 11-point scale with five anchors that ranged from 0, which indicated no pain, to 10, which indicated extreme pain (i.e. 'I can't do anything when I have a headache'). Diary recordings were obtained during the 1-month baseline, during the 2-month treatment/dose adjustment phase and at the 3-month evaluation. The headache index, the mean of all diary ratings (including 0s) for a 1-month period, was the primary outcome measure for the trial (31). The headache index provided a measure of overall headache activity, taking into account both severity and duration. Construct validity of daily diary measures is supported by the association of diary data with independent ratings of impairment by significant others (38), by examination of the discriminate and convergent validity of diary measures (39) and by expert consensus evidence assessments concluding that diary measures provide the most accurate available measure of headache activity (40–42).

The two treating neurologists did not have access to participants' daily headache diary, which was mailed directly to research staff each week. Treating neurologists conducted an independent evaluation of headache activity at each clinic visit, including the number of headache days per 30-day period and pain severity. At the 3-month evaluation as part of a computer-administered assessment, all participants also completed an assessment of headache activity including improvements in headache days and pain severity. Headache improvement ratings were obtained from the neurologist assessment and from participants' report for comparison with daily diary recordings (i.e. headache index).

Side-effects—At each clinic visit treating neurologists coded all side-effects and rated medication side-effects on the following 4-point scale: 0 = none; 1 = mild; 2 = moderate; 3 = severe.

Blindness assessment—As part of a computer-administered questionnaire at the 3-month evaluation, trial participants were asked 'What is your best guess about the preventive medication that you are taking?'. Treating neurologists also were asked 'What is your best

guess about the preventive medication that this patient is taking?'. In both cases possible responses were 'Inactive medication' or 'Active medication'.

Statistics

χ^2 and Fisher's exact tests were used to examine differences in proportions, while *t*-tests for independent samples were used to evaluate differences on continuous variables (e.g. headache index).

Results

Penetration of the blind for participants receiving medication only

Neurologist judgements about the participant's medication condition (AM or PL) were unavailable for five (three AM and two PL) participants. Thus, analyses using this variable included 45 AM and 36 PL participants. Participant judgement of their own medication condition was unavailable for one (AM) participant. Thus analyses using this latter variable included 47 AM and 38 PL participants.

Neurologists correctly identified the medication condition of 82% (66/81) of participants receiving AM or PL alone, a rate that significantly exceeded chance ($\chi^2 = 32.11$, $P = 0.000$). Moreover, neurologists were about equally accurate in identifying the medication condition of participants receiving AM and participants receiving PL ($\chi^2 = 3.68$, $P = 0.06$), correctly identifying the medication condition of 89% (40/45) of participants in the AM group ($\chi^2 = 27.22$, $P = 0.000$) and 72% (26/36) of participants in the PL group ($\chi^2 = 7.11$, $P < 0.008$).

In contrast, only 60% (51/85) of participants receiving AM or PL alone correctly identified their own medication condition, about the number that would be expected by chance ($\chi^2 = 3.40$, $P = 0.07$). However, participants receiving AM were significantly more accurate in identifying their treatment condition than were participants receiving PL ($\chi^2 = 12.07$, $P = 0.001$): 77% (36/47) of participants receiving AM correctly identified their treatment condition ($\chi^2 = 13.30$, $P = 0.000$), but only 40% (15/38) of participants receiving PL correctly identified their treatment condition ($\chi^2 = 1.68$, $P = 0.194$).

Neurologists and trial participants showed moderate agreement on the participants' treatment condition ($\kappa = 0.36$, $P < 0.000$) (Table 1).

Correlates of penetration of the blind for participants receiving medication only

Table 1 compares headache improvement ratings and side-effect ratings when trial participants' treatment conditions were either correctly or incorrectly identified by neurologists and by participants.

Neurologists—First consider the trial participants on AM alone. Participants (correctly) identified by neurologists as receiving AM were rated as more improved than were participants (incorrectly) identified as receiving PL ($P = 0.002$). For participants on PL alone the reverse was true: participants (correctly) identified by neurologists as receiving PL were rated as less improved than were participants (incorrectly) identified as receiving AM ($P = 0.03$). Neurologists thus appear to have identified participants who they perceived to be more improved as being on AM and participants they perceived to be relatively unimproved as being on PL.

Trial participants whom the neurologists (correctly) identified as receiving PL were also rated as being free of side-effects, while participants (incorrectly) identified as receiving AM were rated, on average, as having mild side-effects ($P = 0.003$). However, participants on AM whom

neurologists (correctly) identified as receiving AM were not rated as exhibiting significantly more side-effects than participants (incorrectly) identified as receiving PL. This suggests that neurologists relied primarily on improvement to infer participants' treatment conditions, but may also have used the absence of side-effects to identify trial participants on PL.

Trial participants—Participants, like neurologists, were likely to judge they were on AM if they were improved and judge they were on PL if they were unimproved. Thus, participants who (correctly) identified themselves as receiving AM rated their headaches more improved than participants who (incorrectly) identified themselves as receiving PL ($P = 0.002$). Conversely, participants who (correctly) identified themselves as receiving PL rated their headaches less improved than did participants who (incorrectly) identified themselves as receiving AM ($P = 0.000$). Thus, participants appear to have inferred they were in the AM condition if they perceived their headaches to be improved and in the PL condition if their headaches did not improve. Side-effect ratings did not differentiate participants who correctly and incorrectly identified their respective medication conditions, providing no evidence that participants used the presence of side-effects to infer their treatment condition.

Concordance among outcome measures for participants receiving medication only

All three measures of headache improvement (i.e. neurologists' ratings, participants' ratings and headache index) indicated that participants who received AM only were significantly more improved than participants who received PL only (see Table 2). Moreover, examination of confidence intervals indicates that the AM treatment effect was similar in magnitude when calculated from neurologists' ratings, participants' ratings, or from daily diaries (headache index).

Influence of stress management therapy on the penetration of the medication blind

To examine whether the addition of SMT had any influence on penetration of the medication blind, the identical set of analyses were performed for participants who received medication plus SMT. Neurologist judgements about the participant's medication condition were unavailable for four (PL + SMT) participants. Thus, analyses using this variable included 45 AM + SMT and 34 PL + SMT participants. One participant's judgement (PL + SMT) of her own medication condition was unavailable. Thus, analyses using this latter variable included 45 AM + SMT and 37 PL + SMT participants.

Neurologists' overall accuracy in identifying the medication condition of participants receiving either AM or PL and concurrent SMT (73%; 58/79) was lower, but not significantly so, when compared with their accuracy in identifying the medication condition of participants receiving AM or PL alone (82%; 66/81). However, examining the two individual SMT conditions revealed that neurologists correctly identified the medication condition for 82% (37/45) of participants in the AM + SMT group ($\chi^2 = 18.69$, $P = 0.000$), but did so for only 62% (21/34) of participants in the PL + SMT group ($\chi^2 = 1.88$, $P = 0.17$), thus remaining blind in the latter condition. Thus, while neurologists penetrated the medication blind in the PL-only condition, they were unable to do so in the PL + SMT condition, suggesting that concurrent SMT may have obscured cues to participants' medication condition. However, any such effect of SMT was insufficient to prevent neurologists from penetrating the medication blind in the AM + SMT group.

Trial participants failed to identify their overall medication condition at better than chance levels when they received concurrent SMT (55%; 45/82); moreover, participants in each of the SMT conditions appeared to remain blind, with only 58% (26/45) of participants in the AM + SMT group ($\chi^2 = 1.09$, $P = 0.30$) and 51% (19/37) of participants in the PL + SMT group ($\chi^2 = 0.027$, $P = 0.87$) correctly identifying their own treatment condition. Thus, participants

receiving concurrent SMT, unlike participants receiving medication alone, appeared to be blind as to their medication condition.

Neurologists and trial participants again showed moderate agreement when identifying participants' treatment condition ($\kappa = 0.36$, $P = 0.001$).

Correlates of penetration of the blind in participants receiving concurrent SMT

Penetration of the blind by neurologists appeared to be related to neurologists' ratings of participant improvement, but not to side-effects (Table 3). Participants in the AM + SMT group whom neurologists (correctly) identified as receiving AM were rated more improved than participants who were (incorrectly) identified as receiving PL ($P = 0.003$). Conversely, participants in the PL + SMT group whom neurologists (correctly) identified as receiving PL were rated less improved than participants who were (incorrectly) identified as receiving AM ($P = 0.005$). Neurologists' side-effect ratings did not differ in any comparison.

Influence of concurrent SMT on concordance among outcome measures

All three measures of headache improvement (i.e. neurologists' ratings, participants' ratings and headache index) indicated that participants who received AM + SMT were significantly more improved than participants who received PL + SMT (see Table 4). Moreover, examination of confidence intervals indicates that the AM + SMT treatment effect was similar in magnitude when calculated from neurologists' ratings, participants' ratings or from daily diaries (headache index). The results for participants receiving concurrent SMT were thus consistent with results for the participants treated with medication only.

Discussion

For trial participants receiving medication alone neurologists successfully identified the medication condition for 82% (89% receiving active medication and 72% receiving placebo) of participants at month 3 of treatment. Even in participants receiving concurrent SMT, neurologists correctly identified the medication condition for 82% of participants receiving active medication, although they failed to identify participants receiving placebo at better than chance levels. Trial participants accurately identified their own medication condition when receiving active medication alone (77% of participants), but not when receiving placebo. However, when receiving concurrent SMT participants were unable to identify their medication condition at above chance levels, and thus appeared blind. Our findings for the two medication-only arms of the trial largely replicate those obtained in placebo-controlled trials of psychotropic medications for psychiatric disorders, where often 70% or more of trial participants and treating physicians or evaluators are observed to guess participants' medication condition correctly (6–8,10,13,17–23). Our results are also roughly comparable to findings from a recent trial of amitriptyline vs. 'active' (benztropine mesylate) placebo for spinal cord injury pain (16). In that trial, 62% of trial participants (vs. 60% of participants in the present trial) correctly identified their medication condition and the medication condition of 74% of trial participants was also correctly identified by a research nurse (vs. the medication condition of 84% of participants being correctly identified by project neurologists in the present trial). However, penetration of the blind by the research nurse in the Turner et al. study is noteworthy because the use of an active placebo would be expected to limit the value of side-effects as an indicator of medication condition, and the absence of an amitriptyline treatment effect in the trial would be expected to limit the value of improvement as a cue to medication condition. However, with data from a single research nurse, the generalizability of this finding is uncertain. Taken together, the above studies suggest that penetration of the blind is commonplace, not only in controlled clinical trials in psychiatry, but also in headache and chronic pain trials.

Clinical outcome appeared to be the primary correlate of penetration of the blind. Thus, both neurologists and trial participants appeared to ‘guess’ active medication when the participant improved and placebo when the participant was unimproved. For neurologists, but not trial participants, penetration of the blind in the placebo arm of the trial was also associated with side-effects. Thus, neurologists were more likely to identify placebo correctly when no side-effects were reported. However, the reverse was not true: the correct identification of trial participants on active medication was not associated with the presence of side-effects. Nonetheless, because side-effects are likely to be present early in treatment before improvement occurs, it is possible that neurologists were unblinded to placebo early in treatment, and thus in a position to bias subsequent evaluations, or differentially influence subsequent improvements in the placebo group. To determine the exact point at which penetration of the blind occurred, penetration of the blind would need to be assessed earlier and at multiple points during treatment. We chose not to do this out of fear that repeated questioning of neurologists and trial participants would, itself, influence penetration of the blind.

Concurrent SMT eliminated penetration of the medication blind by trial participants and reduced, but did not eliminate, penetration of the blind by neurologists. Thus, trial participants receiving concurrent SMT were unable to identify their medication condition at better than chance levels. However, even when participants were receiving concurrent SMT, neurologists still correctly identified participants receiving active medication (82% of participants), although they were now unable to identify participants receiving placebo at greater than chance levels. Results reported by Carroll and colleagues (21) appear consistent with our findings. Although the blindness of prescribing physicians was not assessed in the Carroll et al. study, blinded evaluators were less likely to identify correctly the medication condition (desipramine or placebo) of trial participants who were receiving concurrent psychological treatment (65% of participants) than trial participants who were receiving only a brief psychological control procedure (84% of participants).

It has been argued that penetration of the blind leads prodrug clinicians to bias evaluations of improvement, and thus clinical trial outcomes exhibit a prodrug bias when the blind is penetrated (11,23–29,43). In this literature clinician evaluations have been considered biased if they yield more positive results than evaluations by trial participants (27,30), or if ‘subjective’ and ‘objective’ items from the same evaluation instrument appeared to be differentially associated with penetration of the blind (20,21). Unfortunately, there is no consensus that either of these measures enjoy special status as outcome measures and thus are appropriate criterion measures for assessing bias (11,12). In contrast, there is consensus, clearly embodied in guidelines for clinical trials in headache, that mandates contemporaneous daily diaries for clinical trials (40–42,44). We thus chose daily diary data as our criterion measure, although, as we discuss below, daily diary data are not without their limitations. Despite unblinding of both neurologists and trial participants in the medication arms of the trial, we found that clinician evaluations, reports by trial participants and daily diary data yielded identical conclusions about treatment effects. Our results thus gave no indication that either clinician evaluations or trial participant reports exhibited either a pro- or antidrug bias when daily headache diaries were used as the criterion measure.

Several significant limitations of this study should be kept in mind in interpreting these findings. First, discrepancies between clinician evaluations or trial participant reports and daily diary data can be due to factors other than bias. For example, memory processes that lead to an overestimation of the severity of pain (45) may distort retrospective assessments by both clinicians and trial participants. It is therefore at least theoretically possible that clinical evaluations and participant reports were influenced by a prodrug bias, but that all the other factors that can produce a discrepancy between retrospective and daily diary assessments

combined precisely so as to offset any sign of this bias. The possibility that penetration of the blind influenced actual headache activity or the recording of headaches in a produg fashion also cannot be ruled out in the absence of a truly objective measure of headache activity. Despite these limitations, we would argue that our diary data provide at least as strong a criterion measure as the retrospective clinician ratings and retrospective participant reports used as criterion measures for assessing bias in previous studies.

Our findings are also qualified by at least three other study limitations. Relationships between penetration of the blind, improvement and side-effects cannot be interpreted as causal relationships because these variables were assessed at the same time point. Cell sizes were small (<10) and thus statistical power was low for some comparisons involving incorrectly classified participants, because neurologists and trial participants correctly classified the treatment condition for over 80% of participants in some treatment conditions. Both the present trial and the trial by Turner and colleagues evaluated amitriptyline HCl, so results could well differ in headache trials that evaluate other classes of drugs.

The standard practice of assessing penetration of the blind by asking clinician evaluators and trial participants to 'guess' the participant's treatment condition may overestimate actual penetration of the blind. Even when blind to treatment condition, an evaluator who adopts the strategy of guessing active medication for improved, and placebo for unimproved participants, will appear unblinded if blindness is assessed following improvement in a positive trial (46). An examination of neurologists' perceptions of the blindness assessment also suggests the standard procedure for assessing blindness may overestimate penetration of the blind. Thus one study neurologist (F.J.O.) noted that he makes a concerted effort not to think about participants' treatment condition when participating in clinical trials, so the blindness assessment itself disrupted his strategy for remaining blind. These problems argue for improved methods of assessing blindness.

If blindness is to be assessed early and frequently in a trial, so it can be determined if unblinding occurs prior to, or subsequent to improvement, unobtrusive assessment procedures (47) will be needed that can be administered repeatedly without the assessment itself disturbing the blind. If existing procedures for assessing blindness are retained, the assessment of clinical evaluators' and trial participants' confidence in their guess may provide additional useful information: even when two evaluators both correctly guess the treatment condition of a trial participant, the evaluator who is confident they have broken the blind may present more of a threat to the internal validity of the trial than the evaluator who lacks confidence, and is thus truly just guessing. Recent statistical efforts to combine 'guesses' including 'don't know' responses into a blindness index with a known statistical distribution (48,49), thereby allowing a more precise test of blindness, are also a step forward. However, these measures suffer from the limitation of data provided by 'guesses' and might be improved by incorporating information from confidence ratings.

Because daily diary data are mandated as the primary outcome measure in clinical trials in headache (40–42,44), it might be argued that claims that clinician evaluations or trial participant reports exhibit a produg bias are irrelevant to current clinical trials. However, collection of daily diary data may not be feasible in effectiveness studies and health services research (50) where clinician evaluations or trial participant reports may provide the only available outcome measure (e.g. (51,52)). It is thus important to identify the conditions that lead clinician evaluations and participant reports to be concordant with diary recordings in some trials (53,54) and discordant with diary recordings in other trials (38,55,56), as well as to continue to assess the construct validity of daily diary data (e.g. (38)).

In summary, our findings are consistent with studies conducted in psychiatry indicating unblinding occurs frequently in placebo-controlled trials of drug therapies. They further suggest this problem is reduced but not eliminated in trials evaluating combined psychological and drug therapies. These findings suggest that the integrity of the blind in future trials in headache needs to be evaluated rather than assumed. This will require improved methods of assessing and quantifying blindness. However, penetration of the blind does not automatically imply clinician evaluators exhibit a prodrug bias, as some have implied. Despite unblinding, we found no indication that clinician evaluations or reports by trial participants exhibited either a pro- or antidrug bias, at least when daily diary recordings were used as the criterion measure.

Acknowledgements

Support for this study was provided by a grant from the National Institutes of Health (NINDS no. NS32374). Appreciation is expressed to staff involved in conducting the original clinical trial: Athens, OH Site: Bruce Carlson PhD, Katherine Davis-Rosenbaum PhD, Kimberly Hill PhD, Jennifer Labus PhD, Pete Malinoski PhD, Carol Nogrady BA, Cornelia Pinnell PhD and Michael Stensland PhD; Columbus, OH site: Douglas French PhD, Gay Lipchik PhD, Adriana Meade, Angela Nicolosi MA, France Talbot PhD and Sharon Waller RN.

References

- Schulz K, Grimes D. Epidemiology series: blinding in randomized trials: hiding who got what. *Lancet* 2002;359(9307):696–700. [PubMed: 11879884]
- Gracely R, Dunbar R, Deeter W, Wolskee P. Clinician's expectations influence placebo analgesia. *Lancet* 1985;1(8419):43. [PubMed: 2856960]
- Schulz K, Chalmers L, Hayes R, Altman D. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12. [PubMed: 7823387]
- Swartzman L, Burkell J. Expectations and the placebo effect in clinical drug trials: why we should not turn a blind eye to unblinding, and other cautionary notes. *Clin Pharmacol Ther* 1998;64:1–7.
- Kirsch, I., editor. How expectancies shape experience. American Psychological Association; Washington DC: 1999.
- Brownell K, Stunkard A. The double-blind in danger: untoward consequences of informed consent. *Am J Psychiatry* 1982;139:1487–9. [PubMed: 6753613]
- White K, Kando J, Park T, Waternaux C, Brown WA. Side effects and the 'blindability' of clinical drug trials. *Am J Psychiatry* 1992;149:1730–1. [PubMed: 1443253]
- Bystritsky A, Waikar S. Inert placebo versus active medication: patient blindability in clinical pharmacological trials. *J Nerv Ment Dis* 1994;182:485–7. [PubMed: 8083675]
- Thomson R. Side effects and placebo amplification. *Br J Psychiatry* 1982;140:64–8. [PubMed: 7037102]
- Basoglu M, Marks I, Livanou M, Swinson R. Double-blindness procedures, rater blindness, and ratings of outcome. *Arch Gen Psychiatr* 1997;54:744–8. [PubMed: 9283510]
- Greenberg R, Bornstein R, Mj ZSF, Greenberg M. A meta-analysis of fluoxetine outcome in the treatment of depression. *J Nerv Ment Dis* 1994;182:547–51. [PubMed: 7931201]
- Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *Br J Psychiatry* 1998;172:227–31. [PubMed: 9614471]
- Rabkin J, Markowitz J, Stewart J, McGrath P, Harrison W, Quitkin F, et al. How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenazine. *Psychiatry Res* 1986;19:75–86. [PubMed: 3538107]
- Quitkin F, Rabkin J, Gerald J, Davis J, Klein D. Validity of clinical trials of antidepressants. *Am J Psychiatry* 2000;157:327–37. [PubMed: 10698806]
- Montori VM, Bhandari M, Devereaux PJ, Mannis BJ, Ghali WA, Guyatt GH. The reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 2002;55:787–90. [PubMed: 12384193]
- Turner J, Jensen M, Warms C, Cardenas D. Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial. *Pain* 2002;99:91–9. [PubMed: 12237187]

17. Bakker A, Spinhoven P, van Balkom A, Matser D, van Dyk R. Double-blindness procedure did not mask giving of medication in panic disorder. *J Affect Dis* 1999;54:189–92. [PubMed: 10403163]
18. Hughes J, Krahn D. Blindness and the validity of the double-blind procedure. *J Clin Psychopharmacol* 1985;5:138–42. [PubMed: 3998203]
19. Morin C, Colecchi C, Brink D, Astruc M, Mercer J, Remsberg S. How 'blind' are double-blind placebo-controlled trials of benzodiazopine hypnotics? *Sleep* 1995;18:240–5. [PubMed: 7618021]
20. Margraf J, Ehlers A, Roth W, Clark D. How 'blind' are double-blind studies? *J Consult Clin Psychol* 1991;59:184–7. [PubMed: 2002136]
21. Carroll K, Rounsaville B, Nich C. Blind man's bluff: effectiveness and significance of psychotherapy and pharmacotherapy blinding procedures in a clinical trial. *J Consult Clin Psychol* 1994;62:276–80. [PubMed: 8201064]
22. Marini J, Sheard M, Bridges C, Wagner E. An evaluation of the double-blind design in a study comparing lithium carbonate with placebo. *Acta Psychiatr Scand* 1976;53:343–54. [PubMed: 785950]
23. Fisher S, Greenberg R. How sound is the double-blind design for evaluating psychotropic drugs? *J Nerv Ment Dis* 1993;181:345–50. [PubMed: 8501455]
24. Moncrieff J. Are antidepressants overrated? A review of methodological problems in antidepressant trials. *J Nerv Ment Dis* 2001;189:288–95. [PubMed: 11379971]
25. Antonuccio DO, Danton WG, DeNelsky GY, Greenberg RP, Gordon JS. Raising questions about antidepressants. *Psychother Psychosom* 1999;68:3–14. [PubMed: 9873236]
26. Fisher, S.; Greenberg, R., editors. *The limits of biological treatments for psychological distress: comparisons with psychotherapy and placebo*. Lawrence Erlbaum Associates; Hillsdale: 1989.
27. Greenberg RP, Bornstein R, Greenberg M, Fisher S. A meta-analysis of antidepressant outcome under 'blinder' conditions. *J Consult Clin Psychol* 1992;60:664–9. [PubMed: 1401382]
28. Kirsch I, Moore TJ, Scoboria A, Nicholls S. The emperor's new drugs: an analysis of antidepressant medication data submitted to the US Food and Drug Administration. *Prevention & Treatment* 2002;5:23. Available from: <http://www.journals.apa.org/prevention/volume5/pre0050023a.html> Accessed 15 March 2006
29. Kirsch, I.; Sapirstein, G. Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication.. In: Kirsch, I., editor. *How expectancies shape experience*. American Psychological Association; Washington, DC: 1999.
30. Petcova E, Quitkin F, McGrath P. A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology* 2000;22:559–65. [PubMed: 10788756]
31. Holroyd KA, O'Donnell FJ, Stensland M, Lipchik GL, Cordingley GE, Carlson B. Management of chronic tension-type headache with tricyclic antidepressant medication, stress-management therapy, and their combination: a randomized controlled trial. *JAMA* 2001;285:2208–15. [PubMed: 11325322]
32. Olesen JC. Classification and diagnostic criteria for headache disorders, cranial neuralgias, and facial pain. Headache Classification Committee of the International Headache Society. *Cephalalgia* 1988;8 (Suppl 7):1–96.
33. Silberstein S, Rosenberg J. Multispecialty consensus on diagnosis and treatment of headache. *Neurology* 2000;54:1553–4. [PubMed: 10762491]
34. Mathew, NT.; Bendtsen, L. Prophylactic pharmacotherapy of tension-type headache.. In: Olesen, J.; Tfelt-Hansen, P.; Welch, KMA., editors. *The headaches*. 2nd edn. Lippincott Williams & Wilkins; Philadelphia: 2000. p. 667-74.
35. Ashina A, Ashina M. Current and potential future drug therapies for tension-type headache. *Curr Pain Headache Rep* 2003;7:466–74. [PubMed: 14604506]
36. Holroyd, KA.; French, DJ.; Nash, JM.; Tobin, DL.; Echelberger-McCune, RL. Stress management for tension headaches: a treatment program for controlling headaches. Ohio University Headache Project; Athens, OH: 1995.
37. Lipchik, GL.; Holroyd, KA.; Nash, JM. Cognitive-behavioral management of recurrent headache disorders: a minimal-therapist contact approach.. In: Turk, DC.; Gatchel, RS., editors. *Psychological approaches to pain management*. 2nd edn. Guilford Pubs; New York: 2002. p. 356-89.

38. Blanchard EB, Andrasik F, Neff DF, Jurish SE, O'Keefe DM. Social validation of the headache diary. *Behav Ther* 1981;12:711–5.
39. Holroyd KA, Malinoski P, Davis MK, Lipchik GL. The three dimensions of headache impact: pain, disability and affective distress. *Pain* 1999;83:571–8. [PubMed: 10568866]
40. Penzien D, Andrasik F, Freidenberg B, Houle T, Lake A, Lipchik G, et al. Guidelines for trials of behavioral treatments for recurrent headache, first edition. American Headache Society Behavioral Clinical Trials Workgroup. *Headache* 2005;45(Suppl 2):S110–S132. [PubMed: 15921503]
41. Schoenen J, Boureau F, Kunkel R. Guidelines for trials of drug treatments in tension-type headache. *Cephalalgia* 1995;15:165–79.
42. International Headache Society Clinical Trials Subcommittee. Guidelines for Controlled Trials of Drugs in Migraine, second edition. *Cephalalgia* 2000;20:765–86.
43. Greenberg, R.; Fisher, S. A second opinion: rethinking the claims of biological psychiatry.. In: Fisher, S.; Greenberg, R., editors. *The limits of biological treatments for psychological distress: comparisons with psychotherapy and placebo*. Lawrence Erlbaum Associates; Hillsdale: 1989. p. 309-36.
44. Andrasik F, Lipchick G, McCrory D, Wittrock D. Outcome measurement in behavioral headache research: headache parameters and psychosocial outcomes. *Headache* 2005;45:429–37. [PubMed: 15953259]
45. Erskine A, Morley S, Pearce S. Memory for pain: a review. *Pain* 1990;41:255–65. [PubMed: 1697054]
46. Sharpe L, Ryan B, Allard S, Sensky T. Testing for the integrity of blinding in clinical trials: how valid are forced choice paradigms? *Psychother Psychosom* 2003;72:128–31. [PubMed: 12707479]
47. Webb, E.; Campbell, D.; Schwartz, D.; Sechrest, L. *Unobtrusive measures: nonreactive research in social sciences*. Rand McNally; Chicago: 1966.
48. Bang H, Ni L, Davis C. Assessment of blinding in clinical trials. *Control Clin Trials* 2004;25:143–6. [PubMed: 15020033]
49. James K, Lee K, Kraemer H, Fuller R. An index for assessing blindness in a multicenter clinical trial: disulfiram for alcohol cessation—a VA cooperative study. *Stat Med* 1990;15:1421–34. [PubMed: 8841652]
50. Nash J, McCrory D, Nicholson R, Andrasik F. Efficacy and effectiveness approaches in behavioral treatment trials. *Headache* 2005;45:507–12. [PubMed: 15953267]
51. Rothrock J, Parada V, Sims C, Key K, Walters N, Zweifler R. The impact of intensive patient education on clinical outcome in a clinic-based migraine population. *Headache* 2005;45:772–3.
52. Maizels M, Saenz V, Wirjo J. Impact of a group-based model of disease management for headache. *Headache* 2003;43:621–7. [PubMed: 12786921]
53. Stewart WF, Lipton RB, Kolodner K, Sawyer J, Lee C, Liberman J. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain* 2000;88:41–52. [PubMed: 11098098]
54. Holroyd KA, Nash JM, Pingel JD, Cordingley GE, Jerome A. A comparison of pharmacological (amitriptyline HCl) and nonpharmacological (cognitive-behavioral) therapies for chronic tension headaches. *J Consult Clin Psychol* 1991;59:387–93. [PubMed: 2071723]
55. Holroyd KA, Penzien DB. Pharmacological vs. nonpharmacological prophylaxis of recurrent migraine headache: a meta-analytic review of clinical trials. *Pain* 1990;42:1–13. [PubMed: 2146583]
56. Cahn T, Cram J. Changing measurement instrument at follow-up: a potential source of error. *Biofeedback Self Regul* 1980;5:265–73. [PubMed: 7397279]

Comparison of clinical improvement and side-effect ratings when neurologists and trial participants correctly or incorrectly identified the participant's medication condition*

Treatment group Ratings	Active medication Classification ('guess')		t(df) [†]	P [‡]	Placebo Classification ('guess')		t(d.f)	P
	Correct	Incorrect			Correct	Incorrect		
Neurologists' ratings								
N	40	5	—	—	26	10	—	—
Improvement [§]	44.7 (25.6)	6.7 (17.2)	3.2 (1.43)	0.002	6.5 (14.2)	30 (49.9)	-2.2 (1.34)	0.033
Side-effects [¶]	0.8 (0.8)	0.2 (0.5)	1.8 (1.43)	0.076	0.00 (0.00)	0.6 (1.0)	-3.2 (1.34)	0.003
Trial participants' ratings								
N	36	11	—	—	15	23	—	—
Improvement	57.8 (26.6)	28.2 (26.9)	3.2 (1.45)	0.002	1.6 (27.3)	38.1 (25.6)	-3.9 (1.36)	0.000
Side-effects	0.8 (0.8)	0.5 (0.7)	1.4 (1.45)	0.173	0.1 (0.3)	0.2 (0.7)	-0.8 (1.36)	0.413

* Neurologist and trial participant 'guesses' either correctly or incorrectly classified each participant as receiving active medication or placebo.

[†] t-value for comparison.

[‡] P-value for comparison.

[§] Neurologist and participant global assessments are expressed as mean (SD) percentage improvement.

[¶] Mean (SD) side-effect ratings on a 0–3 severity scale.

Table 2
Effect of antidepressant medication as assessed by neurologist ratings, trial participant ratings, and daily diaries (headache index) *

	Active medication	Placebo	t(df) [†]	P [‡]	Effect size	95% CI [§]
N [¶]	48	38	—	—	—	—
Neurologist improvement rating	39.5 (27)	14.1 (30.1)	4.1 (1.84)	0.000	25.4	(13.1, 37.7)
Participant improvement rating	50.9 (29.2)	23.7 (33.1)	4.0 (1.83)	0.000	27.2	(13.7, 40.6)
Headache index improvement	46.2 (32.3)	13.2 (49)	3.6 (1.84)	0.001	32.9	(14.5, 51.6)

* Neurologists' ratings, trial participants' ratings and diary recordings (headache index) expressed as mean (SD) percentage improvement.

[†] t-value with degrees of freedom for comparison.

[‡] P-value for comparison

[§] CI, Confidence interval.

[¶] One (AM) participant improvement rating was unavailable.

Comparison of clinical improvement and side-effects ratings when neurologists and trial participants correctly or incorrectly identified the participant's medication condition *

Table 3

Ratings	Active medication + stress-management therapy Classification ('guess')			t(df) [†]	P [‡]	Placebo + stress-management therapy Classification ('guess')			t(df)	P
	Correct	Incorrect				Correct	Incorrect			
Neurologists N [§]										
Improvement [¶]	37	8		—	—	21	13		—	—
Side-effects rating ^{**}	50.6 (24.7)	19.6 (25.6)		3.2 (1.42)	0.003	15.6 (20.1)	40.1 (26.1)		-3 (1.31)	0.005
Trial participants N [§]	0.7 (0.8)	0.2 (0.5)		1.4 (1.43)	0.159	0.1 (0.2)	0.1 (0.3)		-0.3 (1.31)	0.761
Improvement [¶]	26	19		—	—	19	18		—	—
Side-effects rating ^{**}	69.9 (19.7)	45.1 (21.9)		4 (1.43)	0.000	39.8 (26.3)	52.6 (39.2)		-1.2 (1.35)	0.257
	0.7 (0.8)	0.4 (0.6)		1.7 (1.43)	0.097	0 (0)	0.1 (0.3)		-1.5 (1.34)	0.163

* Neurologist or trial participant 'guesses' either correctly or incorrectly classified each participant as receiving active medication or placebo.
[†] t-value for comparison.
[‡] P-value for comparison.
[§] Neurologist improvement rating missing for one (AM + SMT) participant; neurologist improvement and side-effect ratings missing for one (PL + SMT) participant.
[¶] Neurologist and participant global assessments expressed as mean (SD) percentage improvement.
^{**} Mean (SD) side effect ratings on a 0–3 severity scale.
^{††} Neurologist side-effect rating missing for one (PL + SMT) participant.

Effect of antidepressant medication plus stress-management therapy as assessed by neurologist ratings, trial participant ratings, and daily diaries (headache index)*

Table 4

	Active medication + SMT	Placebo + SMT	t(df) [†]	P [‡]	Effect size	95% CI [§]
N [¶]	45	38	—	—	—	—
Neurologist improvement rating	45 (27.4)	26.6 (27.7)	3.0 (1.79)	0.003	18.4	(6.3, 30.6)
Participant improvement rating	59.4 (23.8)	46 (33.4)	2.1 (1.80)	0.045	13.4	(0.3, 26.4)
Headache index improvement	52.7 (28.6)	31.3 (39.9)	2.8 (1.81)	0.006	21.4	(6.4, 36.4)

SMT, Stress-management therapy.

* Neurologists' ratings, trial participants' ratings and daily diary recordings (headache index) expressed as mean (SD) percentage improvement.

[†] t-value with degrees of freedom for comparison.

[‡] P-value for comparison.

[§] CI, Confidence interval.

[¶] One neurologist improvement rating unavailable for each group; one participant improvement rating (PL + SMT) unavailable.