Active placebos versus antidepressants for depression (Review)

Moncrieff J, Wessely S, Hardy R

THE COCHRANE COLLABORATION®

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2009, Issue 1

http://www.thecochranelibrary.com

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

TABLE OF CONTENTS

HEADER . . . 1
ABSTRACT . . . 1
PLAIN LANGUAGE SUMMARY . . . 2
BACKGROUND . . . 3
OBJECTIVES . . . 3
METHODS . . . 3
RESULTS . . . 5
DISCUSSION . . . 7
AUTHORS' CONCLUSIONS . . . 9
ACKNOWLEDGEMENTS . . . 9
REFERENCES . . . 9
CHARACTERISTICS OF STUDIES . . . 11
DATA AND ANALYSES . . . 17
Analysis 1.1. Comparison 1 antidepressant versus active placebo (all trials), Outcome 1 change in mood after treatment period. . . . 19
Analysis 2.1. Comparison 2 subgroup analysis: in patients, Outcome 1 change in mood after treatment. . . . 20
Analysis 3.1. Comparison 3 subgroup analysis: out patients, Outcome 1 change in mood after treatment. . . . 20
Analysis 4.1. Comparison 4 sensitivity analysis (excluding Daneman 1961), Outcome 1 change in mood after treatment. 21
Analysis 5.1. Comparison 5 sensitivity analysis (excluding Daneman 1961 and Murphy et al 1984), Outcome 1 change in mood after treatment. . . . 22
Analysis 6.1. Comparison 6 sensitivity analysis (excluding Daneman 1961 and using higher effect size from Weintraub & Aronson, 1963), Outcome 1 change in mood after treatment. . . . 23
Analysis 7.1. Comparison 7 sensitivity analysis, inpatients (using higher effect size from Weintraub & Aronson, 1963), Outcome 1 change in mood after treatment. . . . 24
Analysis 8.1. Comparison 8 sensitivity analysis, outpatients (excluding Daneman 1961), Outcome 1 change in mood after treatment. . . . 24
Analysis 9.1. Comparison 9 sensitivty analysis (excluding trials with categorical outcomes), Outcome 1 change in mood after treatment. . . . 25
Analysis 10.1. Comparison 10 sensitivity analysis (excluding trials with estimated s.d.'s), Outcome 1 change in mood. 26
WHAT'S NEW . . . 26
HISTORY . . . 26
CONTRIBUTIONS OF AUTHORS . . . 26
DECLARATIONS OF INTEREST . . . 27
SOURCES OF SUPPORT . . . 27
INDEX TERMS . . . 27

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

[Intervention Review]

Active placebos versus antidepressants for depression

Joanna Moncrieff[1], Simon Wessely[2], Rebecca Hardy[3]

[1]University College London, Warley Hospital, Essex, UK. [2]Academic Dept of Psychological Medicine, Guy's, King's, & St Thomas, School of Medicine & Institute of Psychiatry, 103 Denmark Hill, London, UK. [3]Department of Epidemiology and Public Health, MRC National Survey of Health and Development, London, UK

Contact address: Joanna Moncrieff, University College London, Warley Hospital, Mascalls Lane, Brentwood, Essex, CM14 4TU, UK. j.moncrieff@ucl.ac.uk. (Editorial group: Cochrane Depression, Anxiety and Neurosis Group.)

Cochrane Database of Systematic Reviews, Issue 1, 2009 (Status in this issue: *Edited*)
Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.
DOI: 10.1002/14651858.CD003012.pub2
This version first published online: 26 January 2004 in Issue 1, 2004. Re-published online with edits: 21 January 2009 in Issue 1, 2009.
Last assessed as up-to-date: 11 February 2008. (Help document - Dates and Statuses explained)

This record should be cited as: Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database of Systematic Reviews* 2004, Issue 1. Art. No.: CD003012. DOI: 10.1002/14651858.CD003012.pub2.

ABSTRACT

Background

Although there is a consensus that antidepressants are effective in depression, placebo effects are also thought to be substantial. Side effects of antidepressants may reveal the identity of medication to participants or investigators and thus may bias the results of conventional trials using inert placebos. Using an 'active' placebo which mimics some of the side effects of antidepressants may help to counteract this potential bias.

Objectives

To investigate the efficacy of antidepressants when compared with 'active' placebos.

Search strategy

CCDANCTR-Studies and CCDANCTR-References were searched on 12/2/2008. Reference lists from relavant articles and textbooks were searched.

Selection criteria

Randomised and quasi randomised controlled trials comparing antidepressants with active placebos in people with depression.

Data collection and analysis

Since many different outcome measures were used a standard measure of effect was calculated for each trial. A subgroup analysis of inpatient and outpatient trials was conducted. Two reviewers independently assessed whether each trial met inclusion criteria.

Main results

Nine studies involving 751 partcipants were included. Two of them produced effect sizes which showed a consistent and statistically significant difference in favour of the active drug. Combining all studies produced a pooled estimate of effect of 0.39 standard deviations (confidence interval, 0.24 to 0.54) in favour of the antidepressant measured by improvement in mood. There was high heterogeneity due to one strongly positive trial. Sensitivity analysis omitting this trial reduced the pooled effect to 0.17 (0.00 to 0.34). The pooled effect for inpatient and outpatient trials was highly sensitive to decisions about which combination of data was included but inpatient trials produced the lowest effects.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Authors' conclusions

The more conservative estimates from the present analysis found that differences between antidepressants and active placebos were small. This suggests that unblinding effects may inflate the efficacy of antidepressants in trials using inert placebos. Further research into unblinding is warranted.

PLAIN LANGUAGE SUMMARY

Tricyclic antidepressants compared with active placebos for depression

This review examined trials which compared antidepressants with 'active' placebos, that is placebos containing active substances which mimic side effects of antidepressants. Small differences were found in favour of antidepressants in terms of improvements in mood. This suggests that the effects of antidepressants may generally be overestimated and their placebo effects may be underestimated.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

BACKGROUND

Since the 1970's there has been a consensus, based predominantly on the results of clinical trials, that tricyclic antidepressants (TCAs) have a specific therapeutic effect in depression. However, an examination of the literature reveals that the evidence from such trials is not consistent. Although most reviews find that the drug is significantly superior to a placebo in a majority cases, the degree of superiority is generally not large and between 22% and 73% of studies or comparisons fail to find a significant difference (Cole 1964; Davis 1965; Klerman 1967; McNair 1974; Morris 1974; Rogers 1975). In addition, comprehensive analyses of early antidepressant research revealed that the methodology employed influenced the result. In particular, the absence of random allocation or blinding increased the apparent effect size (Smith 1969, Wechsler 1965), a finding which has been noted more recently in other areas of medicine (Schulz 1995, Schulz 1996).

A further methodological concern is the possibility of bias due to unblinding effects. Greenberg 1994a have pointed out that the different physiological experiences resulting from the ingestion of an active drug and an inert placebo may lead patients and assessors to suspect the identify of medication. This may introduce bias due to different expectations of treatment. Several studies have found that drugs, including antidepressants, can be distinguished from placebo more readily than would be predicted by chance (White 1992). There are various possible explanations for this unblinding effect and its possible association with outcome. Unblinding may occur due to the therapeutic effect of the medication or may occur due to side effects but correlate with therapeutic effect. In both these circumstances the therapeutic effect would determine the outcome and how it was measured. However, another suggestion is that side effects may enhance the placebo effect in patients taking active medication (Thomson 1982). A further possibility is that the occurrence of side effects may unblind raters which may produce biased ratings. In these latter situations, outcome may be determined by factors other than the specific effect of the medication, that is results may be biased.

There is some evidence that unblinding effects may be associated with outcome ratings in the absence of evidence that the drug is effective. A drug trial with problem drinkers found that perception of medication group predicted outcome rating, although there was no evidence the drug was effective (Toneatto 1992) and similar findings were reported in a trial of antipsychotic drugs (Engelhardt 1969). In addition side effects have been shown to correlate significantly with patient and clinician outcome depression ratings in a meta-analysis of placebo controlled trials of fluoxetine (Greenberg 1994b).

Some investigators have addressed this difficulty by using placebos containing active substances. Small doses of drugs with anticholinergic actions have typically been employed to mimic side effects of TCAs in placebo preparations. Thomson 1982 reviewed some of these studies and found that they were more likely to have a negative outcome than studies using inert placebos. It would be difficult to conduct a trial using an active placebo at present because many clinicians would feel it was unethical. Meta-analysis of previous active placebo controlled trials therefore provides an opportunity to investigate the efficacy of antidepressants under conditions of greater "blindness". In addition, by combining results of several small trials with various groups of depressed patients meta-analysis should increase the power to detect an effect and balance the idiosyncracies of individual trials.

Depression is the commonest psychiatric condition and antidepressant drug treatment accounted for 1.9% of all NHS drug costs in the early 1990s (Henry 1993). This proportion is likely to have increased since then due to the escalation in prescribing of the new SSRI drugs (Donoghue 1996). Antidepressants are lucrative products for the pharmaceutical industry which therefore devotes much research to the development of new agents. These new drugs are frequently expensive and represent a potentially significant and escalating drain on health service resources unless properly evaluated. Since new drugs have been evaluated by comparison with the gold standard of TCAs the results of this review also have implications for evaluation of the role of newer drugs such as the SSRIs.

OBJECTIVES

To investigate the efficacy of antidepressants when compared with 'active' placebos for treating people with depression. An active placebo is a placebo tablet which contains a drug which is not thought to have a specific effect in the disorder being treated and which is employed to mimic the effect of taking an active substance.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised and quasi randomised trials which were conducted double blind.

Types of participants

Participants of either sex of all age groups whose primary diagnosis was of a depressive disorder. A concurrent diagnosis of another psychiatric or medical disorder was not an exclusion criteria.

Types of interventions

Interventions included any currently used antidepressant drug or antidepressants which have been withdrawn for reasons other than lack of efficacy. To be considered trials also had to use a placebo containing some active substance employed to mimic the non specific effects of taking an active drug.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Types of outcome measures

Trials were included if they used some measurement of depression as an outcome variable. Any type of measure was admissable (since most of the trials found were conducted before the development of outcome measures in current use).

Search methods for identification of studies

1. The following databases were searched with the following strategies
CCDANCTR-Studies (searched on 12/2/2008)
Intervention = "Active Placebo".
CCDANCTR-References (searched on 12/2/2008)
Free text = "active placebo*"

2. Reference lists of relevant papers were scanned for published reports and citations of unpublished research.
3. Book chapters on treatment of depression were scanned for descriptions of trials.

Data collection and analysis

Two reviewers (JM & RH) assessed studies independently to decide whether they met inclusion criteria.

Data extraction
Many different outcomes measures were used, and it was assumed that they all measured an underlying construct which we have called mood. Standardised mean differences (the difference between the group means divided by the combined standard deviation) were used to calculate a standard measure of effect for each trial. Change in mood at the end of treatment was defined as the outcome of interest. In some studies this information was presented directly for the outcome scale used. In one study (Murphy 1984) it was calculated by subtracting pre-treatment scores from post-treatment scores. In this case the standard deviation was estimated from another study using the same outcome measure (Rush 1977). In other cases direct measures of improvement or change were used, such as categorical ratings of improvement or use of measures such as the Global Clinical Improvement Scale (Friedman 1975). Observer rated measures were selected in preference to patient rated ones as these were employed most consistently. Where there was a choice, the measure indicated by the authors as the one of principle importance was selected. If no principle measure was specified, priority was given to instruments that have been widely used and subject to reliability testing, if available data permitted. Where different measures or ratings within the same study disagreed substantially, as occurred in one trial (Weintraub 1963) separate effect sizes were calculated and used in the analysis. Intention to treat data were used where possible and in one trial, with a large number of early withdrawals, this was calculated by assigning the poorest possible outcome to dropouts (Daneman 1961). Results consisting only of categorical ratings of degree of improvement were weighted (e.g. much improved =3, moderate improvement=2, no change=1, worse=0) and mean scores and standard deviations obtained as described in a previous meta-analysis in this area (QAP 1983).
Requests were sent to authors of studies for more complete data and statistics such as standard deviations. However, unsurprisingly none of the data was still available since studies were too old.
Since the number of studies was small, and estimation and approximation was required to produce compatible outcomes, no further analyses were attempted.
Statistical procedures
Standardised mean differences, or 'effect sizes' for the individual trials were calculated by subtracting the mean score in the placebo group from that of the group allocated to antidepressants and dividing the result by the pooled standard deviation. A number of papers did not report standard deviations and so estimates were obtained from other trials using the same outcome measures and similar patient groups. In one study (Uhlenhuth 1964) patients allocated to the antidepressant were more severely depressed at baseline than the placebo group. An effect size adjusted for baseline values was therefore computed using analysis of variance. This adjusted value could not be used in the MetaView analyses but was used for the purpose of correlation analysis with quality scores. Results from individual trials were combined using the MetaView (version 4.0) procedure for standardised mean differences. A fixed effects model was used, because this is the simplest model and there is no consensus as to whether fixed or random effects models are preferable in given situations. In addition, it was felt that heterogeneity between studies should be identified and explored and not incorporated into the effect estimate as would be the case with a random effects model. Heterogeneity was examined visually and statistically.
A subgroup analysis of inpatients and outpatients was planned a priori. Sensitivity analyses were conducted to explore the assumptions made and the consistency of the data. In addition where two or more measures in one trial yielded substantially different outcomes, sensitivity analyses were done using the different effect sizes calculated for each measure.
Quality assessment
There is no consensus on what constitutes quality in randomised controlled trials in psychiatry. Two assessments were conducted for this review. Firstly a qualitative evaluation of the quality of studies was undertaken focusing on allocation, blinding and inclusion of subjects in the analysis. These three aspects of trial design have been found to be the principle determinants of quality in one investigation of trial quality (Jadad 1996). Secondly a more detailed and quantitative examination of trial quality was conducted using an instrument for the assessment of the quality of intervention trials in psychiatry (Moncrieff 2001). This consists of ratings of 23 aspects of trial quality encompassing issues relating to both internal validity or the control of bias, and external validity or generalisability. Each item was scored between zero and two for each

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

trial giving a maximum score of 46.

RESULTS

Description of studies

See: Characteristics of included studies; Characteristics of excluded studies.

The following studies were identified that satisfied all inclusion criteria. Further details are provided in the Table of Characteristics.

Daneman 1961

A parallel group trial of outpatients comparing imipramine with an atropine placebo. It was of variable duration with assessments made at one and two months.

Weintraub 1963

A parallel group study with inpatients comparing imipramine and an atropine placebo over 4 weeks.

Wilson 1963

A factorial study evaluating ECT and imipramine compared with simulated ECT and an atropine placebo with inpatients lasting 5 weeks.

Uhlenhuth 1964

Crossover trial of 4 weeks duration with outpatients for which data is reported as for a parallel group trial at 2 weeks. Compared imipramine with an atropine placebo.

Hollister 1964

Parallel three group trial comparing amitiptyline, imipramine and an atropine placebo in inpatient veterans over 3 weeks.

Friedman 1966

A parallel group trial with inpatients lasting 3 weeks comparing imipramine and an atropine placebo.

Hussain 1970

A parallel three group study of patients from "psychiatric practice" comparing amitriptyline, an amitryptiline and perphenazine combination tablet and an atropine placebo.

Friedman 1975

A parallel group factorial study with married outpatients evaluating marital therapy and amitriptyline using an atropine placebo over 12 weeks.

Murphy, 1984Murphy, 1984

Parallel group trial of cognitive therapy and nortriptylene in outpatients with 12 weeks of treatment. Groups allocated to nortriptylene plus cognitive therapy and cognitive therapy plus active placebo containing atropine and phenobarbital sodium were used in the current analysis.

Three other RCTs comparing antidepressants with active placebos were found. These were not included in the analysis because the subjects were not suffereing from a depressive disorder. Further details are given in the table of characteristics of excluded studies.

Outcome measures:

(See also table of Included studies).

Only two of the trials used the Hamilton Rating Scale for Depression (HRSD) (Wilson 1963, Murphy 1984). Murphy 1984 also used the Beck Depression Inventory. Another study used a modified version but did not report its overall ratings (Friedman 1975). Hollister 1964 used a Manifest Depression scale appended to the Inpatient Multidimensional Psychiatric Inventory (IMPS) which was constructed using ratings of a panel of experts and subjected to factor analysis to explore internal validity (Overall 1962). The authors had used this scale in several previous studies. They also used various scales derived from the Minnesota Multiphasic Personality Inventory (MMPI). Two studies (Friedman 1966, Friedman 1975) used a Global Clinical Improvement Scale which the authors say was described by DiMascio, but no reference was given or could be traced. However, this scale appears to be similar to the much used Clinical Global Impressions Scale which was certainly in widespread use before it was officially described by Guy 1976. Both these studies used several other outcome measures but did not report them fully. Uhlenhuth 1964 describe the development of a scale called the Total Distress Score. This was constructed by rating symptoms from a commonly used symptom checklist (Frank 1957) according to the degree of distress the patient was suffering. Forty two symptoms relevant to the evaluation of depression were then selected via the agreement of a panel of eight senior psychiatrists acting independently. They also used a scale called the Morale Loss Scale derived from the MMPI. Other studies only used or reported ratings of improvement in various numbers of categories.

Risk of bias in included studies

Quality of studies

The simple overview of trial quality revealed some strengths despite the age of most of the studies. Inclusion criteria ensured that they were conducted double blind and had taken measures to strengthen this procedure by using an active placebo. They all used random allocation and although only two did an explicit intention to treat analysis (Friedman 1975; Murphy 1984), all but one (Daneman 1961) of the others documented only small numbers of early withdrawals. Two studies tested the integrity of the blind in assessors by asking for guesses of medication group and although guesses were more accurate than would be predicted by chance, the effect was not statistically significant in either trial (Uhlenhuth 1964; Weintraub 1963). However, in the Weintraub 1963 trial it was found that both raters assessed those they guessed to be on the active drug as more improved. One other trial reported that side effects had been more prominent in patients on antidepressants (Hollister 1964), indicating the possibility that residual unblinding effects may have occurred despite the use of active placebos.

In the more extensive procedure using the quality rating instrument the mean score of the nine studies was 20 (maximum possible score 46, s.d. 6.71). Correlation analysis demonstrated an inverse association between quality score and effect size with a correlation

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

coefficient of -0.605 (p=0.09) and a positive association between quality score and later year of publication (r=0.414, p=0.3). However, the power of correlation analysis was limited by the small number of studies and hence neither of these associations reached statistical significance. Graphical inspection of the relationship between effect size and quality revealed an approximately linear relationship with one outlying study (Daneman 1961). Excluding this study resulted in a correlation coefficient of -0.775 for the association in the eight remaining studies which was statistically significant at the 5% level (p=0.02).

Effects of interventions

Individual studies

Nine trials, involving 751 participants were included. All compared TCAs with active placebos containing atropine. A minimum dose of 100mg of amitriptyline or equivalent was used in all studies except one where the dose used was not stated (Hussain 1970). The effect sizes (SMDs) calculated for each study in units of standard deviation are listed below according to a fixed effects model.

Daneman 1961

This trial showed a positive and significant difference favouring imipramine over active placebo.

SMD = 1.1 (95% confidence interval, C. I., 0.8 to 1.4). Calculated from scored categories of response to treatment. Based on 101 patients allocated to imipramine and 94 to placebo.

Uhlenhuth 1964

This trial showed no difference between imipramine and placebo when the results were adjusted for substantial differences in baseline levels of depression.

Unadjusted SMD = 0.60 (95% C.I. 0.02, 1.2). Calculated on Total Distress Score pre minus post treatment scores (individual patient data was provided and so exact scores could be computed). Based on 22 patients allocated to imipramine and 20 to placebo.

SMD adjusted for baseline values = 0.35 (95% C.I. -0.25 to 0.96). (Not shown in metaview. Calculated using multiple regression analysis).

Weintraub 1963

Results for two different raters were inconsistent with one finding a significant advantage for imipramine over placebo and the other finding no significant difference.

SMD for hospital director = 0.14 (95% C.I. -0.34 to 0.62). Based on 36 patients allocated to imipramine and 31 to placebo.

SMD for ward doctor = 0.63 (95% C.I. 0.15 to 1.11). Based on 36 patients allocated to imipramine and 32 to placebo.

Calculated from scored categories of "improvement"

Wilson 1963

No difference between imipramine and placebo.

Effect size = -0.26 (95% C.I. -1.10 to 0.58). Calculated from change in Hamilton Rating Scale for Depression (HRSD) scores between pre and post treatment measurements. Based on 10 patients allocated to imipramine and 12 to placebo.

Hollister 1964

No difference between two tricyclic antidepressants (imipramine and amitriptyline) and placebo.

SMD = 0.19 (95% C.I. -0.24 to 0.63). Calculated from change in Inpatient Multi-dimensional Psychiatric Scale (IMPS) between pre and post treatment measures. Based on 62 patients allocated to one of the antidepressants and 31 to placebo. Standard deviation estimated from Hollister 1963.

Friedman 1966

No difference between imipramine and placebo.

SMD = 0.13 (95% C.I. -0.37 to 0.64)

Calculated from Global Clinical Improvement scale. Based on 36 patients allocated to imipramine and 26 to placebo. Standard deviation estimated from results at 4 weeks in trial by Friedman 1975.

Hussain 1970

The effect size in this trial indicated that antidepressants were superior to placebo, although the authors found no significant difference using a categorical analysis.

SMD = 0.79 (95% C.I. 0.09 to 1.5)

Calculated from scored categories of improvement.

Based on 15 patients allocated to imipramine and 17 to placebo.

Friedman 1975

No difference between amitriptyline and placebo.

SMD = 0.14 (95% C.I. -0.14 to 0.42).

Calculated from Global Clinical Improvement scale. Based on 98 patients in each group.

Murphy 1984

No difference between nortriptyline and placebo.

Effect size = -0.36 (95% C.I. -1.0 to 0.28)

Calculated from change in HRSD score between pre and post treatment.

Based on 22 patients allocated to nortriptyline and 17 to placebo. Standard deviation estimated from Rush 1977.

Ratings by the two observers in the trial of Weintraub 1963 yielded discrepant estimates of effect size, and pooled meta-analysis was conducted separately using both estimates. In three trials (Friedman 1966; Hollister 1964; Murphy 1984) standard deviations for the relevant measures were not reported and estimates were taken from studies by the same authors or, in one case, from the study that the authors referenced as their blueprint (Rush 1977). Effect sizes calculated in this way were consistent with the results of individual measures reported in the studies and with the authors interpretations of their findings. Two trials showed a consistent and statistically significant difference favouring the antidepressant drug over placebo (Daneman 1961; Hussain 1970), although only one of these authors (Daneman 1961) concluded that an effect had been demonstrated. Adjustment for baseline discrepancies in the severity of depressive symptoms made a marked difference in the trial by Uhlenhuth 1964. Post treatment scores in this study were virtually identical for the intervention and control group, implying that the greater change score in the group

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

allocated to antidepressants may partly represent regression to the mean.

Combined analysis

The distribution of the effect sizes calculated fitted a normal distribution. Tests of skewness and kurtosis were not significant (skewness=0.39, p=0.50; kurtosis 2.19, p=0.89) (Stata). Therefore parametric methods for combining trial statistics could be used.

Combining effect sizes from all nine trials, using the more conservative estimate from Weintraub 1963 (rating by hospital director), yielded a pooled estimate of 0.39 (95% C.I. 0.24 to 0.54). This indicates a highly significant difference between antidepressants and placebos. However, a high degree of heterogeneity was revealed (X2 = 36.3, degrees of freedom, d.f. 8, p<0.001) . Inspection of the results indicated that the source of heterogeneity was likely to be one trial by Daneman 1961, with other results being reasonably consistent. This trial produced a large positive effect size of 1.1 (95% C.I. 0.8 to 1.4) despite assuming a poor outcome in subjects lost to follow up. It yielded an even larger estimate of 2.80 (95% C.I. 2.41 to 3.19) when these assumptions were not made and the improvement rate in the placebo group was unusually poor (9% at eight weeks). Closer inspection revealed the possibility that rating of response was not blind and that selective reporting of outcomes had occurred. It was therefore decided to repeat the analysis excluding this study. This reduced heterogeneity to a non significant level (X2= 8.51, d.f. 7, p=0.29). The pooled effect size for the eight remaining trials was 0.17 (95% C.I. 0.00 to 0.34).

Repeating these analyses with the higher estimate from the trial by Weintraub 1963 marginally increased the size of the overall estimates. In particular it increased the pooled effect for the eight trials excluding Daneman 1961 to 0.23 (95% C.I. 0.06 to 0.40). It did not influence heterogeneity findings.

Excluding the study by Murphy 1984, on the grounds that all participants received cognitive therapy as well as medication, also increased pooled effects a little without affecting heterogeneity. The combined effect size for seven trials excluding Daneman 1961and Murphy 1984 was 0.21 (95% C.I. 0.03, 0.38).

Sensitivity analysis was performed excluding trials in which categorical data was transformed into continuous data. This analysis revealed a low and non signifcant estimate of effect (SMD= 0.13, 95% C.I. -0.06 to 0.31), but it was little different from the estimate of effect obtained by excluding the Daneman 1961 trial alone. Sensitivity analysis was also performed excluding trials in which estimated standard deviations had been used. This produced a higher estimate of effect of 0.51 (0.33, 0.68) based on the six remaining trials and 0.22 (95% C.I. 0.02, 0.43) on the five other trials excluding Daneman 1961.

Inpatient trials predominantly involved patients with endogenous or severe depression. The majority of subjects in outpatient trials were diagnosed as having neurotic or moderate depression. Subgroup analysis in inpatients produced a small and non significant pooled effect size of 0.12 (95% C.I. -0.14 to 0.38) using the lower of the two estimates from Weintraub 1963. Heterogeneity was low and non significant. Using the higher estimate from this trial increased the combined effect to 0.25 (95% C.I. 0.00, 0.51) with no discernable effect on heterogeneity.

Combined analysis with all five outpatient trials produced an effect size of 0.52 (95% C.I. 0.34, 0.70). Again heterogeneity was high (X2=29.1, p<0.001). Excluding Daneman 1961reduced the heterogeneity to a non statistically significant level (X2=7.38, p=0.06) and reduced the effect size to 0.20 (95% C.I. -0.02, 0.43).

DISCUSSION

Limitations of review.

This study demonstrates the difficulty of performing meta-analysis with small numbers of trials because of the sensitivity of the results to the inclusion or exclusion of individual studies. For this reason, decisions about which studies to include in the analysis and which estimates of effect to use should be explicit, and results of sensitivity analyses should be presented. The exclusion of the large trial by Daneman 1961, which was the source of significant heterogeneity, had the most substantial impact on this meta-analysis. It is generally recommended that the source of heterogeneity should be investigated rather than proceeding with a combined analysis of discrepant results (Abramson 1990). In this case it was apparent that the results of this study were inconsistent with the other studies in this review as well as with well known trials using inert placebos (MRC 1965).

In addition, calculating effect size was rarely straight forward, involving conversion of categorical ratings to continuous data and the use of estimated standard deviations in some cases. Sensitivity analysis excluding trials in which categorical data was transformed into continuous data, was little different from the estimate of effect obtained by excluding the Daneman 1961 trial alone. Sensitivity analysis excluding trials in which estimated standard deviations had been used produced slightly higher estimates of effect, since it involved excluding two trials in which antidepressants did not perform better than placebo.

A further problem is that data on change may be skewed and the calculation of effect size is based on parametric statistics. There is no research into how robust these methods are to skewed data. In the trial by Uhlenhuth 1964 in which individual data were available the data did not deviate significantly from the normal distribution (X2 for combined skewness and kurtosis was 3.65, p=0.16) (Stata). However it was apparent that data which had been transformed from categorical ratings were skewed but sensitivity analysis ommitting these trials did not change the results.

Such problems are endemic to meta-analysis in the absence of standard forms of measurement and reporting and are especially prevalent in older trials. They limit the accuracy of the results. How-

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ever, the general interpretation of the results should be consistent with a more qualitative review of the individual studies included. The effect sizes computed were all consistent with indivudal study results and authors' conclusions.

However, the results of a meta-analysis are only as good as the trials on which it is based. Most trials in this review were conducted before operationalised diagnostic criteria were available and when standardised outcome measures were still being developed. The outcome measures used were a mixture which included unvalidated categorical ratings of improvement as well as standardised instruments such as the HRSD and measures developed by the authors of various trials using methods they describe. A global improvement scale similar to the widely used CGI was used in 2 papers. It was necessary to use this mixture of outcome measures in order to use data from all the trials. However, the use of standardised measures is not a panacea. Establishing validity in a condition such as depression is a complex task and existing measures have only been shown to correlate with each other and not with any objective measure of depression. In addition research into the reliability and comparative validity of current measures such as the HRSD has been criticised for using concurrent interviews and inappropriate statistics. When these latter issues are addressed estimates of reliability and validity are much lower (Cicchetti 1983).

The short duration of most of the studies should also be noted, which may make differences between drugs and placebos more difficult to detect. However, all studies used random allocation and by virtue of the inclusion criteria they had all taken measures to strengthen the double blind by use of an active placebo. Also, numbers of exclusions after allocation were small in all but one study. Thus, the studies had all addressed some of the most important aspects of quality whose influence has only recently been widely publicised.

An alternative explanation of the present findings is that atropine itself has antidepressant properties and hence acts not as a placebo in these trials, but as a specific therapeutic agent. Although some open studies have suggested that this may be the case (Kasper 1981), this was not confirmed in a randomised controlled trial comparing centrally and peripherally acting anticholinergic agents which found no difference in their effect on mood (Gillin 1995).

Summary of results.

The limitations of the quantitative analysis and of the individual trials themselves mean that interpretation of results must remain tentative.

All except one of the individual studies were fairly consistent in finding a small, and in most cases non significant, difference between antidepressant drugs and an active atropine placebo. The pooled estimates of effect varied according to which combination of studies was used. The most conservative estimate was 0.17 standard deviations and the least conservative was 0.39. Assuming a normal response to treatment, these estimates indicate that the average score of people taking antidepressant drugs exceeds that of between 57% and 65% of people taking placebo. Alternatively, using the standard deviations reported by Friedman 1975, the estimates would translate into a difference of between 0.4 and 0.8 on the 6 point Clinical Global Improvement Scale. The more conservative estimates might be preferred because of the reasons given for the exclusion of the trial by Daneman 1961, and because the findings about unblinding and rating bias in the trial by Weintraub 1963. In addition, results will have been inflated somewhat because it was not possible to use a measure of effect adjusted for the discrepancy in baseline values in the trial by Uhlenhuth 1964 in the pooled analysis. The large unadjusted effect in this trial may represent partly a regression to the mean effect, since both groups ended the trial with the same levels of depression. There was also evidence of residual unblinding in some of the trials in this review, and the possibility of publication bias may also suggest a more conservative interpretation of results is appropriate. However, the larger estimates of effect are more consistent with other estimates (see below) of the effects of antidepressant drugs.

Subgroup analyses, based on place of care, which was associated with severity of depression, were highly sensitive to decisions about which trials and outcomes to include. The small numbers involved also limited power and accuracy. Conservative estimates showed small and non significant effects in both subgroups.

Quality analysis is in line with previous findings which suggest that poor methodology may inflate the apparent effects of treatment in antidepressant trials (Smith 1969; Wechsler 1965).

Comparisons with other meta-analyses.

Previous meta-analyses of drug treatment of depression have produced diverse estimates of effect size. The largest estimates of 0.81 (95% C.I. 0.65 to 0.97) for endogenous depression and 0.55 (95% C.I. 0.43 to 0.67) for neurotic depression were found in the QAP 1983. Other general samples of trials produced effect sizes of 0.4 (Smith 1980) and 0.67 (Steinbrueck 1983). The smallest estimate came from a review of trials comparing a new antidepressant with both a standard drug and a placebo. It was hypothesised that this design would reduce the influence of expectation on the performance of the standard drug. "Older" antidepressants yielded a combined effect size of 0.25 ($p<0.001$) using observer rated measures and 0.06 (not statistically significant) with patient ratings (Greenberg 1992).

The more conservative estimates from the present study are similar in magnitude to the pooled observer rated outcomes in the review by Greenberg 1992, which would be consistent with the hypothesis that effect sizes in antidepressant trials are inflated by expectations of participants, including researchers. However, confidence intervals were wide and the less conservative estimates, which included the Daneman 1961 trial, were closer to combined results obtained from unselected analyses of antidepressant trials.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

AUTHORS' CONCLUSIONS

Implications for practice

It is difficult to draw firm conclusions from this review because of the small number of trials and the sensitivity of the pooled analysis to inclusion and exclusion of trials with discrepant results.

However, inspection of effect sizes from individual trials revealed that the majority of trials found only small differences between antidepressants and active placebos. Excluding the trial which was the source of heterogeneity resulted in a relatively small pooled effect. It may therefore be the case that unblinding effects have an impact on the results of antidepressant trials using inert placebos and help to inflate the results of other unselective meta-analyses. The specific effects of antidepressants may therefore be smaller than is generally believed, with the placebo effect accounting for more of the clinical improvement observed than is already known to be the case. This would imply that the risks of antidepressant therapy are less likely to be outweighed by their benefits than is currently held to be the case. It might therefore be appropriate to reassess the current pattern of widespread prescribing of antidepressants. However, the age and quality of the studies and the problems of meta-analysis in this situation should not be disregarded and mean that these conclusions must remain tentative.

Implications for research

Further research into unblinding and its impact on antidepressant trials is desirable to clarify this area of concern. Research into safe active placebos may enable further trials with active placebos to be conducted. Given the extent of their current use, it would be particularly interesting to be able to compare the new generation of antidepressants such as the SSRIs to active placbos. In the meantime, testing the integrity of the double-blind in trials using inert placebos provides some idea of the extent to which unblinding occurs. This procedure is recommended for future clinical trials.

ACKNOWLEDGEMENTS

South Thames NHS executive for funding the review as part of a research training fellowship.

Hugh McGuire, CCDAN trials search co-ordinator, for performing the electronic searches.

REFERENCES

References to studies included in this review

Daneman 1961 *{published data only}*
Daneman EA. Imipramine in office management of depressive reactions. *Diseases of the Nervous System* 1961;**22**:213–7.

Friedman 1966 *{published data only}*
Friedman AS, Granick S, Cohen HW, Cowitz B. Imipramine (tofranil) vs. placebo in hospitalised psychotic depressives. *Journal of Psychiatric Research* 1966;**4**(1):13–36.

Friedman 1975 *{published data only}*
Friedman AS. Interaction of drug therapy with marital therapy in depressive patients. *Archives of General Psychiatry* 1975;**32**(5):619–37.

Hollister 1964 *{published data only}*
Hollister LE, Overall JE, Johnson M, Pennington V, Katz G, Shelton J. Controlled comparison of imipramine, amitriptyline and placebo in hospitalised depressed patients. *Journal of Nervous and Mental Disease* 1964;**139**(4):370–5.

Hussain 1970 *{published data only}*
Hussain Z. Drugs in depressive illness. *BMJ* 1970;**1**(707):482.

Murphy 1984 *{published data only}*
Murphy GE, Simons AD, Wetzel RD, Lustman PJ. Cognitive therapy and pharmacotherapy. *Archives of General Psychiatry* 1984;**41**(1):33–41.

Uhlenhuth 1964 *{published data only}*
Uhlenhuth EH, Park LC. The influence of medication (imipramine) and doctor in relieving depressed psychoneurotic outpatients. *Journal of Psychiatric Research* 1964;**2**(2):101–22.

Weintraub 1963 *{published data only}*
Weintraub W, Aronson H. Clinical judgement in psychopharmacological research. *Journal of Neuropsychiatry* 1963;**5**(1):65–70.

Wilson 1963 *{published data only}*
Wilson IC, Vernon JT, Guin T, Sandifer MG. A controlled study of treatments of depression. *Journal of Neuropsychiatry* 1963;**4**:331–7.

References to studies excluded from this review

Azima 1962a *{published data only}*
Azima H, Arthurs D, Silver A, Azima F. The effect of MP-809 in depressive states: a multi-blind study.. *American Journal of Psychiatry* 1962;**119**:573–74.

Azima 1962b *{published data only}*
Azima H, Silver A, Arthurs D. The effects of G-33040 in depressive states: a multi-blind study.. *American Journal of Psychiatry* 1962; **119**:465–6.

Giannini 1986 *{published data only}*
Giannini AJ, Malone DA, Giannini MC, Price WA, Loiselle RH. Treatment of depression in chronic cocaine and phencyclidine abuse with desipramine.. *J Clin Pharmaco* 1986;**26**(3):211–214.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Max 1987 *{published data only}*

Max MB, Culnane M, Schafer SC, Gracely RH, Walther DJ, Smoller B, Dubner R. Amitriptyline relieves diabetic neuropathy pain in patients with normal or depressed mood.. *Neurology* 1989;**37**(4):589–96.

Max 1991 *{published data only}*

Max MB, Kishore Kumar R, Schafer SC, Meister B, Gracely RH, Smoller B, Dubner R. Efficacy of desipramine in painful diabetic neuropathy: a placebo-controlled trial. *Pain* 1991;**45**(1):3–9.

Additional references

Abramson 1990

Abramson JH. Meta-analysis: a review of the pros and cons. *Public Health Review* 1990;**18**(1):1–47.

Cicchetti 1983

Cicchetti DV, Prusoff BA. Reliability of depression and associated clinical symptoms.. *Archives of General Psychiatry* 1983;**40**(9):987–90.

Cole 1964

Cole J. Therapeutic efficacy of antidepressant drugs. *JAMA* 1964; **190**:448–55.

Davis 1965

Davis JM. Efficacy of tranquilising and antidepressant drugs. *Archives of General psychiatry* 1965;**13**(6):552–72.

Donoghue 1996

Donoghue J, Tylee A, Wildgust H. Cross sectional database analysis of antidepressant prescribing in general practice in the United Kingdom, 1993-5. *BMJ* 1996;**313**(7061):861–2.

Engelhardt 1969

Engelhardt DM, Margolis RA, Rudorfer L, Paley HM. Physician bias and the double-blind. *Archives of General Psychiatry* 1969;**20**(3): 315–20.

Frank 1957

Frank JD, Gleidman LH, Imber SD, Nash EH, Stone AR. Why patients leave psychotherapy. *Archives of Neurology and Psychiatry* 1957;**77**:283–95.

Gillin 1995

Gillin JC, Laurellio J, Kelsoe JR, Rapaport M, Golshan S, Kenny WM, et al.No antidepressant effect of biperiden compared with placebo in depression: A double-blind 6-week clinical trial. *Psychiatric Research* 1995;**58**(2):99–105.

Greenberg 1992

Greenberg RP, Bornstein RF, Greenberg MD, Fisher S. A meta-analysis of antidepressant outcome under "blinder" conditions. *Journal of Consulting and Clinical Psychology* 1992;**60**(5):664–9.

Greenberg 1994a

Greenberg RP, Fisher S. Suspended judgement. Seeing through the double-masked design: a commentary. *Controlled Clinical Trials* 1994;**15**(4):244–6.

Greenberg 1994b

Greenberg RP, Bornstein RF, Zborowski MJ, Fisher S, Greenberg MD. A meta-analysis of fluoxetine outcome in the treatment of depression. *Journal of Nervous and Mental Disease* 1994;**182**(10):547–51.

Guy 1976

Guy W. *ECDEU Assessments Manual for Psychopharmacology*. Rockville, MD: National Institute for Mental Health, 1976.

Henry 1993

Henry JA. Debits and credits in the management of depression. *British Journal of Psychiatry Supplementum* 1993;**20**:33–39.

Hollister 1963

Hollister LE, Overall JE, Johnson M, Katz G, Kimbell I, Honigfeld G. Evaluation of desipramine in depressive states. *Journal of New Drugs* 1963;**3**:161–6.

Jadad 1996

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al.Assessing the quality of reports of randomised clinical trials: is blinding necessary?. *Controlled Clinical Trials* 1996;**17**(1): 1–12.

Kasper 1981

Kasper S, Moises HW, Beckman H. The anticholinergic biperiden in depressive disorders. *Pharmacopsychiatria* 1981;**14**(6):195–8.

Klerman 1967

Klerman GL, Cole J. Clinical pharmacology of imipramine and related compounds. *International Journal of Psychiatry* 1967;**3**(4):267–304.

McNair 1974

McNair DM. Self-evaluations of antidepressants. *Psychopharmacologia* 1974;**37**(4):281–302.

Moncrieff 2001

Moncrieff J, Churchill R, Drummond DC, McGuire H. Development of a quality assessment instrument for trials of treatments for depression and neurosis. *International Journal of Methods in Psychiatric Research* 2001;**10**(3):126–33.

Morris 1974

Morris JB, Beck AT. The efficacy of antidepressant drugs. A review of reaearch (1958-1972). *Archives of General Psychiatry* 1974;**30**(5): 667–74.

MRC 1965

Medical Research Council. Clinical trial of the treatment of depressive illness. *BMJ* 1965;**1**:881–6.

Murray 1989

Murray EJ. Measurement issues in the evaluation of psychopharmacological therapy. In: Fisher S, Greenberg RP editor(s). *The limits of biological treatments for psychological distress*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989:39–68.

Overall 1962

Overall JE. Dimensions of manifest depression. *Journal of Psychiatric Research* 1962;**1**:239–47.

QAP 1983

The Quality Assurance Project. A treatment outline for depressive disorder. *Australia and New Zealand Journal of Psychiatry* 1983;**17**: 129–46.

Rogers 1975

Rogers SC, Clay PM. A statistical review of controlled trials of imipramine and placebo in the treatment of depressive illness. *British Journal of Psychiatry* 1975;**127**:599–603.

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Rush 1977
Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative effects of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Research and Therapy* 1977;**1**(1):17–37.

Schulz 1995
Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**(5):408–412.

Schulz 1996
Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of parallel group trials in obstetrics and gynaecology.. *BMJ* 1996;**312**(7033):742–4.

Smith 1969
Smith A, Traganza E, Harrison G. Studies on the effectiveness of antidepressants drugs. *Psychopharmacology* 1969;**Suppl**:1–53.

Smith 1980
Smith MC, Glass GV, Miller TI. Reviews of research on the effects of drug therapy on psychological disorders. In: Smith MC, Glass GV, Miller TI editor(s). *The benefits of psychotherapy*. Baltimore, MD: John Hopkins University Press, 1980:127–261.

Stata
Stata Press. Stata. 5. College Station, TX: Stata Press, 1997.

Steinbrueck 1983
Steinbrueck SM, Maxwell SE, Howard GS. A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *Journal of Consulting and Clinical Psychology* 1983;**51**(6):856–63.

Thomson 1982
Thomson R. Side effects and placebo amplification. *British Journal of Psychiatry* 1982;**140**:64–68.

Toneatto 1992
Toneatto T, Sellers EM. Perception of whether drug or placebo has been administered is a determinant of drinking reduction. In: Sellers EM, Naranjo CA editor(s). *Novel pharmacological interventions for alcoholism*. New York, NY: Springer Verlag, 1992.

Wechsler 1965
Wechsler H, Grosser GH, Greenblatt M. Research evaluating antidepressants in hospitalised mental patients: a survey of published reports during a five year period. *Journal of Nervous and Mental Disease* 1965;**141**(2):231–9.

White 1992
White K, Kando J, Park T, Waternaux C, Brown WA. Side effects and the "blindability" of clinical drug trials. *American Journal of Psychiatry* 1992;**149**(12):1730–61.

References to other published versions of this review

Moncrieff 1998
Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *British Journal of Psychiatry* 1998;**172**:227–31.

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Daneman 1961

Methods	parallel group trial. Variable duration with evaultions done at 1 month and 2 months.
Participants	195 outpatients, age range 17-75, 69% women
Interventions	imipramine mean dose 133mg and atropine 1.25 mg
Outcomes	4 "response to treatment" categories
Notes	not clear if response to treatment, which was based on ratings of a list of symptoms, was rated blind.
Risk of bias	

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Daneman 1961 (*Continued*)

Item	Authors' judgement	Description
Allocation concealment?	Yes	A - Adequate

Friedman 1966

Methods	parallel group trial. Duration 3 weeks.
Participants	78 inpatients
Interventions	imipramine 150-200mg placebo contained atropine (dose not reported)
Outcomes	Global Clinical Improvement on 6 point scale rated by project psychiatrist and ward doctor, Philadelphia Psychiatric Centre Psychatric Rating Scale (30 items), Philadelphia Psychiatric Center Depression Progress Test, Clyde Mood Scale plus psychometric tests.
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Friedman 1975

Methods	parallel group factorial trial evaluating marital therapy and amitriptyline. Duration 12 weeks.
Participants	196 married outpatients, mean age 36, range 21-67; 79% women
Interventions	amitriptyline 100mg placebo contained atropine 0.4mg
Outcomes	Global Clinical Improvement Scale (score 1-6), Psychiatric Rating Scale (based on HRSD), Patient Self Report Inventory of Psychic and Somatic Complaints, family role, marital relations
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Hollister 1964

Methods	parallel group trial comparing imipramine, amitriptyline and placebo. Duration 3 weeks
Participants	110 inpatient in veterans hospitals, median age 43, range 26-72; all men
Interventions	imipramine mean dose 171mg, amitriptyline mean dose 157mg placebo contained atropine 1mg
Outcomes	5 subscales from Inpateint Multidimensional Psychiatric Scale: manifest depression, anxious intropunitiveness, retardation, conceptual disorganisation, excitement; 2 subscales from Minnesota Multiphasic Personality Inventory: manifest depression scale and "D" scale
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Hussain 1970

Methods	parallel group trial comparing amitriptyline, amitiptyline + perphenazine and placebo. Duration not reported.
Participants	34 patients from psychiatric practice, no details reported
Interventions	doses not reported placebo contained atropine (dose not reported)
Outcomes	5 categories of improvement
Notes	This is a brief communication about preliminary results in a letter. No final report of this trial could be traced.

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Murphy 1984

Methods	parallel group cognitive therapy trial. Groups had adjunctive cognitive therapy. Duration 12 weeks treatment followed by 4 weeks follow up.
Participants	39 outpatients involved in this comparison, age range of completers 19-59, 66% completers women
Interventions	nortriptyline 100-150mg placebo contained atropine 0.1-0.15mg and phenobarbital sodium 10-15mg
Outcomes	Hamilton Rating Scale for Depression, Beck Depression Inventory, Scale for Suicidal Ideation, Hopelessness Scale, Raskin, Three-Area Severity of Depression Scale, Visual Analogue Scale, Zung Anxiety Scale, Social Adjustment Scale, MMPI, Self Control Scale, Cognitive Response Test, Dsysfunctional Attitude Scale, Automatic Thoughts Questionnaire
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Yes	A - Adequate

Uhlenhuth 1964

Methods	crossover trial of 4 weeks with results of first period of 2 weeks reported as for parallel groups.
Participants	50 outpatients, mean age 42 (range 22-71); 76% women
Interventions	imipramine 150mg atropine 0.6mg
Outcomes	Total Distress Score, Morale Loss Scale, doctors and patients overall estimate of condition as better, same, worse.
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Weintraub 1963

Methods	parallel group study. Duration 4 weeks.
Participants	89 inpatients, 60% women, mean age 51 (range 19-73)
Interventions	imipramine 150 mg atropine 0.6mg
Outcomes	improvement rated in three catgeories by ward doctor and hospital director
Notes	discrepant ratings with ward doctor rating drug group as more improved and finding greater drug placebo difference. Blind tested. Neither rater guessed medication group better than chance but both raters rated those they guessed to be on the drug as significantly more improved (p<0.1).

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Wilson 1963

Methods	factorial design testing ECT vs simulated ECT and drug vs placebo Duration 5 weeks.
Participants	24 inpatients, all women, age range 40-59
Interventions	imipramine 150-220 mg atropine (dose not reported)
Outcomes	Hamilton Rating Scale for Depression, MMPI "D" scale
Notes	

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Unclear	B - Unclear

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Characteristics of excluded studies *[ordered by study ID]*

Azima 1962a	Did not use a recognised antidepressant. Not all observers were blind to treatment allocation.
Azima 1962b	Did not use a recognised antidepressant. Not all observers were blind to treatment allocation.
Giannini 1986	Subjects did not have a diagnosed depressive disorder. (Trial of desipramine for depressive symptoms associated with cocaine and PCP withdrawal.)
Max 1987	Subjects did not have a diagnosed depressive disorder. (Crossover trial of amitiptyline in diabetic neuropathy)
Max 1991	Subjects did not have a diagnosed depressive disorder. (Crossover trial of amitiptyline in diabetic neuropathy)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

DATA AND ANALYSES

Comparison 1. antidepressant versus active placebo (all trials)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment period	9	750	Std. Mean Difference (IV, Fixed, 95% CI)	0.39 [0.24, 0.54]

Comparison 2. subgroup analysis: in patients

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	4	244	Std. Mean Difference (IV, Fixed, 95% CI)	0.12 [-0.14, 0.38]

Comparison 3. subgroup analysis: out patients

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	5	506	Std. Mean Difference (IV, Fixed, 95% CI)	0.52 [0.34, 0.70]

Comparison 4. sensitivity analysis (excluding Daneman 1961)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	8	555	Std. Mean Difference (IV, Fixed, 95% CI)	0.17 [-0.00, 0.34]

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Comparison 5. sensitivity analysis (excluding Daneman 1961 and Murphy et al 1984)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	7	516	Std. Mean Difference (IV, Fixed, 95% CI)	0.21 [0.03, 0.38]

Comparison 6. sensitivity analysis (excluding Daneman 1961 and using higher effect size from Weintraub & Aronson, 1963)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	8	556	Std. Mean Difference (IV, Fixed, 95% CI)	0.23 [0.06, 0.40]

Comparison 7. sensitivity analysis, inpatients (using higher effect size from Weintraub & Aronson, 1963)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	4	245	Std. Mean Difference (IV, Fixed, 95% CI)	0.25 [-0.00, 0.51]

Comparison 8. sensitivity analysis, outpatients (excluding Daneman 1961)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	4	311	Std. Mean Difference (IV, Fixed, 95% CI)	0.20 [-0.02, 0.43]

Comparison 9. sensitivty analysis (excluding trials with categorical outcomes)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood after treatment	6	454	Std. Mean Difference (IV, Fixed, 95% CI)	0.13 [-0.06, 0.31]

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Comparison 10. sensitivity analysis (excluding trials with estimated s.d.'s)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 change in mood	6	556	Std. Mean Difference (IV, Fixed, 95% CI)	0.51 [0.33, 0.68]

Analysis 1.1. Comparison 1 antidepressant versus active placebo (all trials), Outcome 1 change in mood after treatment period.

Review: Active placebos versus antidepressants for depression

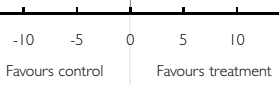
Comparison: 1 antidepressant versus active placebo (all trials)

Outcome: 1 change in mood after treatment period

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Daneman 1961	94	1.64 (1.45)	101	0.37 (0.78)		24.0 %	1.10 [0.80, 1.40]
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		8.6 %	0.13 [-0.37, 0.64]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		27.8 %	0.14 [-0.14, 0.42]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		11.7 %	0.19 [-0.24, 0.63]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		4.4 %	0.79 [0.09, 1.50]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		5.4 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		5.7 %	0.60 [-0.02, 1.22]
Weintraub 1963	36	1.17 (0.7)	31	1.07 (0.68)		9.4 %	0.14 [-0.34, 0.62]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		3.1 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**395**		**355**			**100.0 %**	**0.39 [0.24, 0.54]**

Heterogeneity: Chi^2 = 36.26, df = 8 (P = 0.00002); I^2 =78%

Test for overall effect: Z = 5.18 (P < 0.00001)



Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.1. Comparison 2 subgroup analysis: in patients, Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

Comparison: 2 subgroup analysis: in patients

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		26.1 %	0.13 [-0.37, 0.64]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		35.7 %	0.19 [-0.24, 0.63]
Weintraub 1963	36	1.17 (0.7)	31	1.07 (0.68)		28.8 %	0.14 [-0.34, 0.62]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		9.4 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**144**		**100**			**100.0 %**	**0.12 [-0.14, 0.38]**

Heterogeneity: Chi^2 = 0.90, df = 3 (P = 0.83); I^2 =0.0%

Test for overall effect: Z = 0.91 (P = 0.36)

Analysis 3.1. Comparison 3 subgroup analysis: out patients, Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

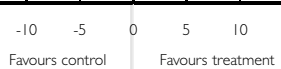
Comparison: 3 subgroup analysis: out patients

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Daneman 1961	94	1.64 (1.45)	101	0.37 (0.78)		35.7 %	1.10 [0.80, 1.40]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		41.4 %	0.14 [-0.14, 0.42]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		6.5 %	0.79 [0.09, 1.50]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		8.0 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		8.5 %	0.60 [-0.02, 1.22]
Total (95% CI)	**251**		**255**			**100.0 %**	**0.52 [0.34, 0.70]**

Heterogeneity: Chi^2 = 29.07, df = 4 (P<0.00001); I^2 =86%

Test for overall effect: Z = 5.68 (P < 0.00001)



Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 4.1. Comparison 4 sensitivity analysis (excluding Daneman 1961), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

Comparison: 4 sensitivity analysis (excluding Daneman 1961)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		11.3 %	0.13 [-0.37, 0.64]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		36.6 %	0.14 [-0.14, 0.42]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		15.4 %	0.19 [-0.24, 0.63]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		5.8 %	0.79 [0.09, 1.50]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		7.1 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		7.5 %	0.60 [-0.02, 1.22]
Weintraub 1963	36	1.17 (0.7)	31	1.07 (0.68)		12.4 %	0.14 [-0.34, 0.62]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		4.0 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**301**		**254**			**100.0 %**	**0.17 [0.00, 0.34]**

Heterogeneity: Chi^2 = 8.51, df = 7 (P = 0.29); I^2 =18%

Test for overall effect: Z = 1.94 (P = 0.053)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 5.1. Comparison 5 sensitivity analysis (excluding Daneman 1961 and Murphy et al 1984), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

Comparison: 5 sensitivity analysis (excluding Daneman 1961 and Murphy et al 1984)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		12.1 %	0.13 [-0.37, 0.64]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		39.3 %	0.14 [-0.14, 0.42]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		16.6 %	0.19 [-0.24, 0.63]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		6.2 %	0.79 [0.09, 1.50]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		8.0 %	0.60 [-0.02, 1.22]
Weintraub 1963	36	1.17 (0.7)	31	1.07 (0.68)		13.4 %	0.14 [-0.34, 0.62]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		4.3 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**279**		**237**			**100.0 %**	**0.21 [0.03, 0.38]**

Heterogeneity: $Chi^2 = 5.71$, df = 6 (P = 0.46); $I^2 = 0.0\%$

Test for overall effect: Z = 2.31 (P = 0.021)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 6.1. Comparison 6 sensitivity analysis (excluding Daneman 1961 and using higher effect size from Weintraub & Aronson, 1963), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

Comparison: 6 sensitivity analysis (excluding Daneman 1961 and using higher effect size from Weintraub % Aronson, 1963)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		11.3 %	0.13 [-0.37, 0.64]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		36.7 %	0.14 [-0.14, 0.42]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		15.5 %	0.19 [-0.24, 0.63]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		5.8 %	0.79 [0.09, 1.50]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		7.1 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		7.5 %	0.60 [-0.02, 1.22]
Weintraub 1963	36	1.42 (0.77)	32	0.94 (0.76)		12.1 %	0.62 [0.13, 1.11]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		4.1 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**301**		**255**			**100.0 %**	**0.23 [0.06, 0.40]**

Heterogeneity: Chi^2 = 11.36, df = 7 (P = 0.12); I^2 =38%

Test for overall effect: Z = 2.60 (P = 0.0093)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 7.1. Comparison 7 sensitivity analysis, inpatients (using higher effect size from Weintraub & Aronson, 1963), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

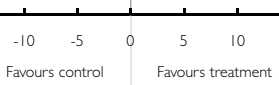
Comparison: 7 sensitivity analysis, inpatients (using higher effect size from Weintraub % Aronson, 1963)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		26.4 %	0.13 [-0.37, 0.64]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		36.0 %	0.19 [-0.24, 0.63]
Weintraub 1963	36	1.42 (0.77)	32	0.94 (0.76)		28.2 %	0.62 [0.13, 1.11]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		9.4 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**144**		**101**			**100.0 %**	**0.25 [0.00, 0.51]**

Heterogeneity: Chi² = 3.89, df = 3 (P = 0.27); I² =23%

Test for overall effect: Z = 1.92 (P = 0.054)



Analysis 8.1. Comparison 8 sensitivity analysis, outpatients (excluding Daneman 1961), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

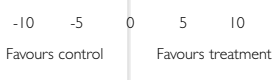
Comparison: 8 sensitivity analysis, outpatients (excluding Daneman 1961)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		64.3 %	0.14 [-0.14, 0.42]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		10.1 %	0.79 [0.09, 1.50]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		12.4 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		13.1 %	0.60 [-0.02, 1.22]
Total (95% CI)	**157**		**154**			**100.0 %**	**0.20 [-0.02, 0.43]**

Heterogeneity: Chi² = 7.38, df = 3 (P = 0.06); I² =59%

Test for overall effect: Z = 1.78 (P = 0.076)



Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 9.1. Comparison 9 sensitivty analysis (excluding trials with categorical outcomes), Outcome 1 change in mood after treatment.

Review: Active placebos versus antidepressants for depression

Comparison: 9 sensitivty analysis (excluding trials with categorical outcomes)

Outcome: 1 change in mood after treatment

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Friedman 1966	36	2.9 (1.9)	26	2.65 (1.84)		13.8 %	0.13 [-0.37, 0.64]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		44.7 %	0.14 [-0.14, 0.42]
Hollister 1964	62	50.17 (57.4)	31	39 (57.4)		18.8 %	0.19 [-0.24, 0.63]
Murphy 1984	22	12.22 (9.22)	17	15.59 (9.22)		8.6 %	-0.36 [-1.00, 0.28]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		9.1 %	0.60 [-0.02, 1.22]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		4.9 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**250**		**204**			**100.0 %**	**0.13 [-0.06, 0.31]**

Heterogeneity: $Chi^2 = 5.32$, df = 5 (P = 0.38); $I^2 = 6\%$

Test for overall effect: Z = 1.33 (P = 0.18)

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 10.1. Comparison 10 sensitivity analysis (excluding trials with estimated s.d.'s), Outcome 1 change in mood.

Review: Active placebos versus antidepressants for depression

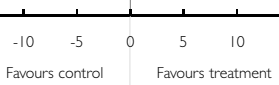
Comparison: 10 sensitivity analysis (excluding trials with estimated s.d.'s)

Outcome: 1 change in mood

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Daneman 1961	94	1.64 (1.45)	101	0.37 (0.78)		32.3 %	1.10 [0.80, 1.40]
Friedman 1975	98	1.85 (2.17)	98	1.54 (2.27)		37.4 %	0.14 [-0.14, 0.42]
Hussain 1970	15	3 (1.6)	19	1.79 (1.4)		5.9 %	0.79 [0.09, 1.50]
Uhlenhuth 1964	22	28.36 (29.05)	20	12.35 (23.06)		7.6 %	0.60 [-0.02, 1.22]
Weintraub 1963	36	1.17 (0.7)	31	1.07 (0.68)		12.7 %	0.14 [-0.34, 0.62]
Wilson 1963	10	15.3 (3.74)	12	16.4 (4.32)		4.1 %	-0.26 [-1.10, 0.58]
Total (95% CI)	**275**		**281**			**100.0 %**	**0.51 [0.33, 0.68]**

Heterogeneity: Chi² = 27.42, df = 5 (P = 0.00005); I² =82%

Test for overall effect: Z = 5.78 (P < 0.00001)



WHAT'S NEW

Last assessed as up-to-date: 11 February 2008.

31 October 2008	Amended	Converted to new review format.

HISTORY

Protocol first published: Issue 2, 1998

Review first published: Issue 2, 2001

6 October 2003	New citation required and conclusions have changed	Substantive amendment

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

CONTRIBUTIONS OF AUTHORS

JM devised protocol, conducted review, wrote review.

RH helped devise protocol, applied inclusion criteria and provided statistical advice.

SW helped devise protocol and provided general supervision for the review project.

DECLARATIONS OF INTEREST

none

SOURCES OF SUPPORT

Internal sources

- South Thames Regional Health Authority, UK.

External sources

- No sources of support supplied

INDEX TERMS

Medical Subject Headings (MeSH)

Antidepressive Agents [*therapeutic use]; Bias (Epidemiology); Depression [*drug therapy]; Placebo Effect; Randomized Controlled Trials as Topic; Treatment Outcome

MeSH check words

Humans

Copyright © 2009 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.