Active placebos versus antidepressants for depression
Moncrieff J, Wessely S, Hardy R


A substantive amendment to this systematic review was last made on 27 July 1998. Cochrane reviews are regularly checked and updated if necessary.

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: The Cochrane Collaboration Depression, Anxiety and Neurosis review group's search strategy was used to search MEDLINE (1966-2000), PsychLIT (1980-2000) and EMBASE (1974-2000) and this was last done in July 2000. Reference lists from relevant articles and textbooks were searched and 12 specialist journals were handsearched up to 1996.

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 participants 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.

Reviewers' 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.

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 & Cole 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.

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 intervention
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.

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 strategy for identification of studies
See: Collaborative Review Group search strategy

1. Electronic databases Medline (1966-2000), PSYCHLIT (1980-2000) and Embase (1974-2000) were searched and searches are on going. The Cochrane Collaboration Depression Anxiety and Neurosis review group
(CCDAN) search strategy for controlled trials in this area was used combined with keywords active placebo and atropine. Searches were last performed in July 2000.

2. The results of handsearching of the following specialist journals were scrutinised up until 1996. When antidepressant trials were identified they were inspected to see whether an active placebo had been used.
   - Archives of General Psychiatry
   - American Journal of Psychiatry
   - British Journal of Psychiatry
   - Acta Psychiatrica Scandinavica
   - Journal of Affective Disorders
   - Clinical Pharmacology and Therapeutics
   - International Clinical Psychopharmacology
   - Journal of Behavioural Research and Therapy
   - Psychopharmacology
   - Psychopathology
   - Cognitive Research and Therapy
   - Clinical Psychopharmacology

3. Reference lists of relevant papers were scanned for published reports and citations of unpublished research.

4. Book chapters on treatment of depression were scanned for descriptions of trials.

Methods of the review

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. Observer rated measures were selected as these were employed most consistently. Where there was a choice, the measure indicated by the authors as the one of principal importance was selected. If no principal 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. Categorical data were only available in four trials and so meta-analysis using dichotomous outcomes was not done.

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, 1963) 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 quality assessment. Results from
individual trials were combined using the MetaView (version 4.0) procedure for standardised mean differences. A fixed effects model was used because it was felt that it was important to explore sources of heterogeneity in the results. MetaView automatically tests for the degree of heterogeneity. 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 principal 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, 1999). 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 trial giving a maximum score of 46.

Description of studies
See: Tables of 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, 1963
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 amitriptyline, 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, 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 suffering 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, 1963 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 form the MMPI. Other studies only used or reported ratings of improvement in various numbers of categories.

Methodological quality
See: Table of 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, 1963; 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 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).

Results

List of comparisons

RESULTS

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, 1963
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, 1963. Post treatment scores in this study were virtually identical for the intervention and control group, implying that the greater change score in the group 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 ($X^2 = 36.3$, 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 ($X^2= 8.51$, 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, 1961 and Murphy, 1984 was 0.21 (95% C.I. 0.03, 0.38).

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 discernible 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 ($X^2=29.1$, $p <0.001$). Excluding Daneman, 1961 reduced the heterogeneity to a non statistically significant level ($X^2=7.38$, $p=0.06$) 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 straightforward, involving conversion of categorical ratings to continuous data and the use of estimated standard deviations in some cases. Sensitivity analysis was performed excluding trials in which categorical data was transformed into continuous data. Because one of these trials was the trial by Daneman, 1961, this analysis revealed a low and non significant 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.

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, 1963 in which individual data were available the data did not deviate significantly from the normal distribution ($X^2$ 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 omitting 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 but should not alter their general interpretation which should be consistent with a more qualitative review of the individual studies included. The effect sizes computed were all consistent with individual 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 Weintrab, 1963 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.

Reviewers' 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 placebos. 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.
Potential conflict of interest
none

References

References to studies included in this review

Daneman, 1961 (published data only)
Friedman, 1966 (published data only)

Friedman, 1975 (published data only)

Hollister, 1964 (published data only)

Hussain, 1970 (published data only)

Murphy, 1984 (published data only)

Uhlenhuth, 1963 (published data only)

Weintraub, 1963 (published data only)

Wilson, 1963 (published data only)

* indicates the major publication for the study

Cover sheet

Active placebos versus antidepressants for depression Reviewer(s) Moncrieff J, Wessely S, Hardy R
Contribution of Reviewer(s) 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.
Issue protocol first published Information not available
Issue review first published Information not available
Date of last minor amendment Information not supplied by reviewer
Date of last substantive amendment 27 July 1998
Most recent changes Information not supplied by reviewer
Date new studies sought but none found Information not supplied by reviewer
Date new studies found but not yet included/excluded Information not supplied by reviewer
Date new studies found and included/excluded Information not supplied by reviewer
Date reviewers' conclusions section amended Information not supplied by reviewer
Contact address Dr Joanna Moncrieff
Specialist registrar
Psychiatry
Charing Cross Hospital
Gloucester House
194, Hammersmith road
London
UK
W6 8BS
Telephone: 00 44 20 8846 7750
Facsimile: 00 44 20 8846 7281
E-mail: joannamoncrieff@compuserve.com
Cochrane Library number CD003012
Editorial group Cochrane Depression, Anxiety and Neurosis Group
Editorial group code HM-DEPRESSN

Sources of support
External sources of support to the review
No sources of support supplied

Internal sources of support to the review

No sources of support supplied

Synopsis
Tricyclic antidepressants are only slightly better than active placebos. 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.

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

Tables & Graphs

Metaview graphs are not available for this review
List of comparisons
Table of included studies
Table of excluded studies
Table of ongoing studies

List of comparisons
A list of comparisons is not available for this review

Tables of other data
Tables of other data are not available for this review

Additional tables
Additional tables are not available for this review

Table of included studies
Study Methods Participants Interventions Outcomes Notes Allocation concealment
Daneman, 1961 parallel group trial. Variable duration with evaluations done at 1 month and 2 months. 195 outpatients, age range 17-75, 69% women imipramine mean dose 133mg and atropine 1.25 mg 4 "response to treatment" categories not clear if response to treatment, which was based on ratings of a list of symptoms, was rated blind. A

Friedman, 1966 parallel group trial. Duration 3 weeks. 78 inpatients imipramine 150-200mg placebo contained atropine (dose not reported) Global Clinical Improvement on 6 point scale rated by project psychiatrist and ward doctor, Philadelphia Psychiatric Centre Psychiatric Rating Scale (30 items), Philadelphia Psychiatric Center Depression Progress Test, Clyde Mood Scale plus psychometric tests. B

Friedman, 1975 parallel group factorial trial evaluating marital therapy and amitriptyline. Duration 12 weeks. 196 married outpatients, mean age 36, range 21-67; 79% women amitriptyline 100mg placebo contained atropine 0.4mg 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 B

Hollister, 1964 parallel group trial comparing imipramine, amitriptyline and placebo. Duration 3 weeks 110 inpatient in veterans hospitals, median age 43, range 26-72; all men imipramine mean dose 171mg, amitriptyline mean dose 157mg placebo contained atropine 1mg 5 subscales from Inpateint Multidimensional Psychiatric Scale: manifest depression, anxious introvertiveness, retardation, conceptual disorganisation, excitement; 2 subscales from Minnesota Multiphasic Personality Inventory: manifest depression scale and "D" scale B
Hussain, 1970  parallel group trial comparing amitriptyline, amitriptyline + perphenazine and placebo. Duration not reported. 34 patients from psychiatric practice, no details reported doses not reported placebo contained atropine (dose not reported) 5 categories of improvement This is a brief communication about preliminary results in a letter. No final report of this trial could be traced. B

Murphy, 1984  parallel group cognitive therapy trial. Groups had adjunctive cognitive therapy. Duration 12 weeks treatment followed by 4 weeks follow up. 39 outpatients involved in this comparison, age range of completers 19-59, 66% completers women nortriptyline 100-150mg placebo contained atropine 0.1-0.15mg and phenobarbital sodium 10-15mg 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, Dysfunctional Attitude Scale, Automatic Thoughts Questionnaire A

Uhlenhuth, 1963  crossover trial of 4 weeks with results of first period of 2 weeks reported as for parallel groups. 50 outpatients, mean age 42 (range 22-71); 76% women imipramine 150mg atropine 0.6mg Total Distress Score, Morale Loss Scale, doctors and patients overall estimate of condition as better, same, worse. B

Weintraub, 1963  parallel group study. Duration 4 weeks. 89 inpatients, 60% women, mean age 51 (range 19-73) imipramine 150 mg atropine 0.6mg improvement rated in three categories by ward doctor and hospital director 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). B

Wilson, 1963  factorial design testing ECT vs simulated ECT and drug vs placebo Duration 5 weeks. 24 inpatients, all women, age range 40-59 imipramine 150-220 mg atropine (dose not reported) Hamilton Rating Scale for Depression, MMPI "D" scale B

---

Table of excluded studies
Study Reason for exclusion
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 amitriptyline in diabetic neuropathy)
max 1991  Subjects did not have a diagnosed depressive disorder. (Crossover trial of amitryptiline in diabetic neuropathy.)

---

Table of ongoing studies
A table of ongoing studies is not available for this review