A comparison of antidepressant trials using active and inert placebos

JOANNA MONCRIEFF¹

¹Department of Psychiatry and Behavioural Science, University College London, UK

ABSTRACT *In placebo-controlled antidepressant trials, there are concerns that the double blind may be breached, which may bias results. Some trials have therefore used placebos containing active substances in order to mimic the side effects of antidepressants. This study set out to examine the impact of 'unblinding' by comparing the results of a meta-analysis of a sample of trials using ordinary inert placebos with a meta-analysis of trials using active placebos. An a priori sub-group analysis of trials conducted with inpatients and outpatients was carried out. Quality was assessed using a quality assessment instrument and meta-regression analysis was conducted to explore heterogeneity. Results of meta-analysis did not differ between trials using active placebos and trials using inert ones. However, results were strongly influenced by two large trials. When these were excluded trials using inert placebos showed greater antidepressant-placebo differences than trials using active placebos. There was very high heterogeneity especially among trials using inert placebos. Meta-regression analysis showed the only significant predictor of outcome was whether trials were conducted with inpatients or outpatients. Sub-group analysis showed significant differences between antidepressants and any type of placebo in trials conducted with outpatients but not in inpatient trials. Lack of data, heterogeneity and the influence of large studies limited the analysis. The integrity of the double blind design, its effect on results and the heterogeneity between studies, require further exploration in antidepressant trials.*

Key words: antidepressants, placebo, meta-analysis, double blind

Introduction

Double-blind, randomized controlled trials (RCTs) are designed to reduce the possibility of bias affecting results and are regarded as the most valid form of scientific evidence about medical interventions. However, there have been concerns raised that the double blind design may not be fool proof and that the occurrence of non therapeutic physiological effects in those taking active drugs may break the blind and affect results (Greenberg and Fisher, 1994). There is evidence from antidepressant trials that participants and assessors can identify the nature of the medication, including SSRIs, better than would be predicted by chance (White et al., 1992; Kranzler et al., 1995). It has also been shown that assessors rate people they believe to be on active treatment as doing better than those they guess are on placebo, in the absence of any demonstrable effect (Toneatto and Sellers, 1992; Engelhardt et al., 1969).

The possibility that trials may be 'unblinded' by non-therapeutic effects potentially threatens the credibility of evidence from RCTs, especially where outcomes are subjective and likely to be influenced by expectations. In the past various efforts were made to reduce the likelihood of unblinding by using an 'active' placebo. In some antidepressant trials a placebo using atropine was employed for comparison with tricyclic antidepressants (TCAs) in order to mimic their anticholinergic side effects. A meta-analysis of these trials was conducted to ascertain the efficacy of antidepressants under conditions of greater 'blindness' (Moncrieff et al., 1998; Moncrieff et al., 2001a). The current study set out to evaluate the effect of using an active placebo compared to a standard inert placebo. The hypothesis was that trials using active placebos would result in smaller drug-placebo differences than those using inert placebos.

It is commonly stated that the effects of antidepressants are greater in more severely depressed patients (Anderson et al., 2000). A secondary hypothesis was investigated that the effect of the type of placebo used would be least apparent in trials conducted with inpatients compared with trials of outpatients. Greater differences between drugs and placebo were expected in the former.

Despite the fact that active placebos have not been used in antidepressant trials since the 1980s, the issue of unblinding remains relevant. There is evidence that newer drugs can be distinguished from placebo (Kranzler et al., 1995), and increasing awareness of antidepressants among the general population may increase the expectations associated with taking medication for depression. In addition, differences between antidepressants and ordinary inert placebos in recent trials have been found to be small (Kirsch, 2002). Given the rapidly rising rates of use of these drugs, it is important to consider any methodological issues that might effect the evidence for their efficacy.

Method

A systematic search was undertaken to locate trials of antidepressants using active placebos. Data were extracted and combined as described in previous publications (Moncrieff et al., 1998; Moncrieff et al., 2001a). Inclusion criteria for all studies were that they should be randomized or quasi randomized and should concern the use of a recognized antidepressant drug for the treatment of depression.

Selection of sample of trials using inert placebos

There are such a large number of controlled trials of antidepressants that it was not possible to examine all trials using inert placebos. A sample was therefore selected using the same inclusion criteria applied to trials using active placebos. Two controlled trials using inert placebos were identified for each trial using an active placebo. In order to obtain a sample of trials with roughly the same chronological spread as the trials using active placebos, trials using inert placebos were matched to the approximate date of publication of the trials using active placebos.

Index Medicus (IM) was used as the sampling frame for trials published before 1966 and this was supplemented by the use of the reference list of a comprehensive review of antidepressant trials (Morris

and Beck, 1974). MEDLINE was used for trials published from 1966 onwards.

The sampling process using Index Medicus began by locating the reference to one of the active placebo trials. Starting at this point and working forward alphabetically subsequent references in the section were scanned and all that looked like trials were noted. The articles were then located and from the ones that reported RCTs and satisfied inclusion criteria, the first two that followed the active placebo trial reference were selected. Where there were not enough trials identified in the same year as the active placebo trial, previous and subsequent years were examined starting at the beginning of the relevant section. This sampling frame was supplemented by the reference list of a comprehensive review by Morris and Beck (1974) starting at the beginning of the reference list and selecting the first trials cited for the requisite year.

Selection of comparison trials for the three active placebo studies published on or after 1966 was carried out using MEDLINE. A search was done using the MESH headings relevant to drug treatment of depression (for example: depression-, drug-, therapy). The search was limited to five years either side of the publication date of the active placebo study. This study was located in the search and then the first references listed above and below the active placebo trial that transpired to be RCTs and fulfilled the inclusion criteria were selected as the controls.

Data extraction

A variety of different outcomes measures were used in the trials examined. These were converted to standardized mean differences (SMD) to obtain a standard measure across trials. Standardized mean differences were calculated using the formula for 'Cohen's d' which involves subtracting the mean score in the placebo group from that of the group on the antidepressant and dividing by the pooled standard deviation (Deeks et al., 2001).

Where there was a choice of measure, the one indicated by the authors as the principal measure was selected where possible. However, the availability of the requisite information (such as standard deviation) for the calculation of effect size sometimes determined which measures could be used. Subsidiary measures were only used where they were consistent with other results obtained in the study, particularly the final

outcome. 'Intention to treat' data were used where it was reported or could be calculated. Results consisting only of categorical ratings of degree of improvement were weighted and mean scores and standard deviations were obtained as described in a previous meta-analysis in this area (Quality Assurance Project, 1983). Two reports of trials using active placebos did not report standard deviations and so estimates were obtained from other similar trials.

Statistical analysis

Stata version 7 was used for all the analyses. A fixed effects model was used and statistical tests of heterogeneity were performed. Where heterogeneity was found to be high possible causes were investigated as recommended rather than using a random effects model (Thompson, 2001). The results of meta-analysis with the sample of active placebo controlled trials was compared with that using inert placebos by examining confidence intervals. Meta-regression analysis was also used to investigate the predictive effects of type of placebo used and to investigate other possible causes of heterogeneity. Meta-regression is a weighted regression analysis that uses the study results as the data points.

A priori subgroup analyses looking at trials with inpatients and outpatients separately were also conducted. This was carried out for trials using active placebos, trials using inert placebos and the combined sample of trials. Meta-regression analysis was also used to assess whether the location of care predicted outcome.

Quality rating

A scale developed for the evaluation of the quality of controlled trials in depression and neurosis was used to assess the quality of all trials (Moncrieff et al., 2001b). This instrument consists of 23 items addressing a range of different aspects of methodological quality. Each item can be scored between 0 (poor) and 2 (good) with the maximum score therefore being 46. Meta-regression analysis was performed to investigate associations between quality score and outcome and also to see whether overall quality varied between trials using active and inert placebos. Meta-regression techniques were also used to explore other possible causes of heterogeneity, including individual aspects of quality, in particular size and duration of study, concealment of allocation, and whether an intention

to treat analysis was performed. Year of publication and dose and type of antidepressant used were also investigated to see whether they predicted outcome.

Publication bias

This was examined using Egger's weighted regression analysis, which essentially uses meta-regression to examine whether the variance of a study predicts outcome (Stata 7). If an association is found this shows that more small positive studies exist than small negative studies, which may suggest publication bias.

Results

Nine trials using active placebos were identified. Their characteristics are summarized in Table 1 and are available in more detail elsewhere (Moncrieff et al., 2001a). A sample of 17 trials using inert placebos was selected. Only one placebo-controlled trial could be located which had been conducted within five years either side of the latest trial conducted with an active placebo. Their characteristics and the outcomes used to calculate SMDs are described in Table 1. Standardized mean differences could not be calculated for two of these trials which were crossover trials and did not present any parallel group data.

Comparison of trials using active placebos and trials using inert placebos

The combined analysis for trials using active placebos yielded an overall effect of 0.39 standard deviations (SDs) (confidence interval, C.I., 0.25 to 0.54) based on analysis of a total of 750 participants. A high degree of heterogeneity between trials was revealed ($Q = 36.8$, d.f. 8, $p < 0.001$). The combined analysis for the sample of trials using inert placebos produced an effect of 0.30 SDs (C.I. 0.18 to 0.41). Again heterogeneity was high ($Q = 51.2$, d.f. 13, $p < 0.001$). Contrary to the hypothesis of this study the confidence intervals of the two estimates overlapped considerably and hence there was no evidence of a statistically significant difference between the two sets of trials.

Overall results were strongly influenced by the results of large studies. When the original meta-analysis of studies using active placebos was done, one outlying study was excluded which reduced heterogeneity to a non-significant level (Moncrieff et al., 1998). However, the estimate obtained from the group of active placebo trials excluding this outlier (SMD =

Table 1. Characteristics of sample of trials using inert placebos

Trial	Source of subjects	Number included in analysis	Duration of trial	Outcome measure used to calculate effect size	SMD (confidence interval)
Trials using active placebos					
Daneman, 1961	Outpatients	195	Up to two months	Five categories of response to treatment	1.1 (0.8 to 1.40)
Ulenhuth and Park, 1963	Outpatients	42	Four weeks	Total Distress Score change	0.35 (−0.25 to 0.96) adjusted for baseline differences
Weintraub and Aronson, 1963	Inpatients	68	Four weeks	Three categories of improvement rated by hospital director	0.15 (0.34 to 0.63)
Wilson et al., 1963	Inpatients	22	Five weeks	HDRS ¹ change	−0.27 (−1.11 to 0.57)
Hollister et al., 1964	Inpatients	93	Three weeks	IMPS ² change	0.19 (−0.24 to 0.63)
Friedman et al., 1966	Inpatients	62	Three weeks	Global Clinical Improvement Scale	0.13 (−0.37 to 0.64)
Hussain, 1970	'Hospital practice'	34	Not reported	Five categories of improvement	0.81 (0.11 to 1.52)
Friedman, 1975	Outpatients	196	12 weeks	Global Clinical Improvement Scale	0.14 (−0.14 to 0.42)
Murphy et al., 1984	Outpatients	39	12 weeks	HRSD change	−0.37 (−1.0 to 0.27)
Trials using inert placebos					
Hall and Lonie, 1962	Inpatients	20	Four weeks	Three categories of improvement	0.26 (−0.62 to 1.15)
Abraham et al., 1963	Outpatients	80	Eight weeks	Five point clinical improvement scale at two weeks	0.96 (0.38 to 1.53)
Garry and Leonard, 1963	Long-term inpatients with chronic depression	20	12 weeks	Not possible to calculate	
Ainslie et al., 1965	Outpatients	34	Four weeks	Change in presenting symptom score (36 signs and symptoms rated 0 to 3)	1.09 (0.22 to 1.95)
Brick et al., 1962	Depressed prisoners	78	Seven weeks	Five categories of improvement	1.10 (0.51 to 1.69)
Hollister et al., 1963	Depressed inpatients at veteran hospital	81	Three weeks	IMPS change score	−0.38 (−0.86 to 0.10)
Rickels et al., 1964	Medical and psychiatric patients with depression	44	Four weeks	Not possible to calculate	
Diamond, 1966	Private out- patients	69	Four to six weeks	Three categories of response	1.09 (0.56 to 1.61)
Prange et al., 1967	Psychiatric patients	26	18–21 days	Three categories of improvement	0.53 (−0.26 to 1.31)
Fink, et al., 1965	All patients referred for drug therapy	35 with depression	Six weeks	MSRPP ³ scale	0.46 (−0.41 to 1.33)
Holdway, 1960	In patients; 34 recent admissions and six long-stay patients	40	Two weeks	Three categories of response at final rating	0.06 (−0.56 to 0.68)
Porter, 1970	General practice patients	60	Three weeks	Global assessment at two weeks	0.1 (−0.4 to 0.61)

Continued

Table 1. Continued

Trial	Source of subjects	Number included in analysis	Duration of trial	Outcome measure used to calculate effect size	SMD (confidence interval)
Trials using inert placebos					
Bellack et al., 1981	Psychiatric outpatients and subjects recruited through advertisements	72	12 weeks	Mean change in HRSD score at final assessment	0.52 (0.05 to 0.99)
Master, 1963	Psychiatric outpatients in India	61	12 weeks	HRSD change at three weeks	0.96 (0.43 to 1.50)
NIMH trial ⁴	Depressed patients from 10 public and private hospital services	379 for imipramine/placebo comparison (555 total)	Five weeks treatment, two weeks follow up	Three categories of response	0.01 (−0.17 to 0.19)
McCallum and Meares, 1975	Psychiatric outpatients	36	Three weeks	HRSD percentage change at final assessment	0.64 (−0.07 to 1.35)
Smith 1975	Psychiatric outpatients	90	Six weeks	Four categories of response	0.44 (−0.08 to 0.96)

1 Hamilton Rating Scale for Depression, (Hamilton, 1960).

2 Inpatient multi-dimensional psychiatric scale (Overall, 1962).

3 Multi dimensional scale for rating psychiatric patients (Lorr et al., 1953).

4 The main publication for this trial is Raskin et al (1970), but no data from this paper could be used to calculate SMD. The SMD is calculated from data given in Raskin and Crook (1975).

0.17, C.I. 0.00 to 0.34) was not significantly different from the combined effect obtained from the comparison sample using inert placebos. The sample of trials using inert placebos was substantially influenced by the results of the large, and essentially negative, National Institute of Mental Health (NIMH) study (Raskin et al., 1970; Raskin and Crook, 1975). When this was excluded a larger estimate of effect of antidepressants compared to inert placebos was obtained (fixed effects model: SMD = 0.52, C.I. 0.36 to 0.68). Confidence intervals just failed to overlap with the combined result obtained from the active placebo trials minus the outlier, but the difference was only statistically significant using a fixed rather than a random effects model.

Table 2 shows that using meta-regression analysis did not reveal a statistically significant association between type of placebo used in the 24 trials with calculable SMDs. When the outlying trial by Daneman (1961) and the NIMH trial were excluded the difference was greater and almost reached statistical significance (Beta = −0.37, s.e. of beta 0.19, $z = -1.93$, $p = 0.05$).

Multivariate meta-regression analysis was also conducted by adding inpatient or outpatient status

into the equation, because this was found to have a strong predictive effect on outcome (see below). However, this did not affect the results of the predictive effect of the type of placebo used, which remained non-significant.

Inpatient versus outpatient trials

Subgroup analyses with inpatients and outpatients are shown in Table 3. In trials using both active and inert placebos differences between antidepressants and placebo were small and non-significant in inpatients. In trials conducted with outpatients differences were larger and statistically significant in both sets of trials and confidence intervals of results overlapped. When the sets of trials were combined, meta-analysis of trials with inpatients showed no difference between antidepressants and placebo and heterogeneity was low. Meta-analysis of all outpatients trials revealed a large and significant difference in favour of antidepressants, but there was substantial heterogeneity among trials which was highly statistically significant. Meta-regression analysis confirmed that whether studies were conducted with inpatients or outpatients strongly predicted outcome (Table 2). Sensitivity analysis was conducted by omitting the two large and outlying

Table 2. Meta-regression of predictive effects of methodological attributes of studies, year of publication and drugs used on effect (SMD) (all analyses reported for N = 24 trials)

Variable (all variables are continuous unless specified)	Beta coefficient (s.e.)	Z value	P value
Type of placebo ¹	-0.18 (0.21)	-0.9	0.37
Inpatient versus outpatient care ²	-0.61 (0.170)	-3.65	<0.001
Overall Quality Score	-0.02 (0.02)	1.23	0.22
Size of study (number of participants)	-0.001 (0.001)	-0.63	0.53
Duration of study (weeks)	0.03 (0.03)	0.85	0.40
Year of publication	-0.02 (-0.02)	-1.34	0.18
Intention to Treat analysis ³	0.01 (0.21)	0.07	0.95
Imipramine versus other drugs ⁴	-0.61 (0.21)	-0.29	0.77
Dose of antidepressant (mg) ⁵	-0.002 (0.002)	-1.54	0.12

1 Entered as a binary variable. A negative coefficient represents a possible association between use of active placebo and smaller effect.

2 Entered as a binary variable. A negative coefficient represents a possible association between inpatient care and smaller effects.

3 Entered as a binary variable with trials including 95% or more of subjects in the analysis defined as performing intent-to-treat analysis.

A positive coefficient represents a possible association between trials using an intent to treat analysis and larger effects.

4 Entered as a binary variable comparing trials using imipramine only and trials using other antidepressants, mostly other TCAs.

A negative coefficient represents a possible association between higher effects and use of drugs other than imipramine.

5 Mean dose was used if provided. If only a dose range was given, the minimum dose specified has been used.

studies (Daneman 1961; Raskin and Crook, 1975). This did not change this result substantially.

Quality analysis and investigation of heterogeneity

The Quality Assessment Instrument was applied to all trials. The maximum possible quality score was 46. The mean score for the combined sample of trials was 18.4, SD was 6.0, range 8 to 32. There was no association between overall quality score and outcome as shown by meta-regression analysis (Table 2). This effect was not altered by omitting the small trial by Hussain (1970), which was published only briefly in the form of a letter. This was done on the grounds that the type of publication did not allow many details about study quality to be reported and hence the assessment of its quality is thought to be unreliable (beta = -0.02, s.e. 0.02, $z = -1.0$, $p = 0.31$).

Meta-regression analysis was also used to investigate whether the type of placebo used predicted study quality. Although the direction of the results suggested that trials using active placebos had higher quality scores, the predictive effect was not statistically significant both in the total combined sample of studies (beta = 2.1, s.e. 2.6, $z = 0.8$, $p = 0.41$) and in the combined sample minus the study by Hussain (1970) (beta = 3.6, s.e. 2.5, $z = 1.5$, $p = 0.14$).

Further exploration of the high degree of heterogeneity between study results was performed by investigating the predictive effects of individual aspects of quality and some other factors. Table 2 shows that no associations were found between SMD and overall size of the study, duration of the study, year of publication and whether an 'intention-to-treat' analysis was performed. The predictive effects of method of concealment of allocation could not be investigated because it was not reported in any of the trials used. The type of antidepressant and whether adequate doses were used were also found to have no effect on study results.

Publication bias

For the combined sample of trials, Egger's weighted regression test for publication bias produced a regression coefficient of 1.31, s.e. 0.9, t value was 1.45, $p = 0.16$.

When the sample was divided into trials using active and trials using inert placebos, it was found that Egger's test showed significant publication bias in the sample of 15 trials using inert placebos (regression coefficient 2.32, s.e. 0.85, t value 2.71, $p = 0.02$), but not in the nine trials using active placebos (regression coefficient -1.97, s.e. 2.07, t value -0.95, $p = 0.38$).

Table 3. Meta-analysis of subgroups of inpatient and outpatient trials (using fixed effects methods)

	Inpatient trials			Outpatient trials		
	Number of trials/ number of participants	Combined effect (95% C.I.)	Heterogeneity (Q)	Number of trials/ number of participants	Combined effect (95% C.I.)	Heterogeneity (Q)
Active placebo trials	4/245	0.12 (−0.14 to 0.38)	Q = 0.95, d.f. 3, p = 0.8	5/506	0.53* (0.35 to 0.71)	Q = 29.5, d.f. 4, p < 0.001
Inert placebo trials	4/512	0.02 (−0.18 to 0.14)	Q = 2.8, d.f. 3, p = 0.4	11/548	0.7* (0.52 to 0.88)	Q = 13.7, d.f. 10, p = 0.19
All trials	8/757	0.02 (−0.12 to 0.15)	Q = 4.6, d.f. 7, p = 0.7	16/1054	0.61* (0.49 to 0.74)	Q = 44.9, d.f. 15, p < 0.001

*p < 0.001

Discussion

Limitations of study

Trials using active placebos have not been conducted for many years and hence the trials used in this analysis were mainly published prior to the widespread adoption of standardized diagnostic systems, outcome measures and reporting standards. This means that it was difficult to obtain data that could be re-analysed and combined. Categorical data were transformed into continuous data in some cases and estimated standard deviations were also used. However, in all cases SMDs were consistent with individual study results. Many studies were also small, and of short duration and there was evidence of publication bias among studies using inert placebos.

Combined results were also substantially influenced by a few trials that were much larger than the rest. Although this is not necessarily inappropriate for a straightforward meta-analysis, it complicates the current objectives. The sample of trials using inert placebos was selected to represent the average findings of such trials. However, combined analysis with this sample was strongly influenced by the results of the large NIMH trial. In addition the results of the meta-analysis with active placebo trials were influenced by one large trial whose results were inconsistent with the other trials (Daneman, 1961).

It is also a concern that there may be publication bias among studies using inert placebos suggesting that unpublished negative studies may exist. Alternatively the bias may have resulted from the selection procedure, although it is not obvious how this would have occurred.

It is difficult to extrapolate findings from this sample of trials to modern-day antidepressant trials, which are conducted using standardised diagnostic instruments and outcome measures, and generally involve community patients and newer antidepressants rather than tricyclics. However, recent research shows that differences between antidepressants and placebo in more recent studies may be very small (Kirsch et al., 2002). Hence knowledge about the integrity of the double blind and its effects on results is particularly important. Recommendations have recently been made for the introduction of procedures that will help to evaluate the integrity of the double blind in antidepressant studies (Even et al., 2000).

Comparison of trials using active and inert placebos

The hypothesis that antidepressant trials using active placebos would find smaller effects than trials using inert placebos was not substantiated in this analysis. By excluding the large outlying study using an active placebo and the large study using an inert placebo the results tended in the direction predicted, but as these were *post hoc* exclusions not too much importance should be attached to this finding. Quitkin et al. (2000) have argued that too much emphasis has been placed on results of trials using active placebos. This study might lend support to this conclusion. However, examining data on the integrity of the blind in the trials using active placebos suggested that the use of an active placebo did not entirely prevent identification of medication group (Moncrieff et al., 1998). Some degree of 'unblinding' may therefore still have occurred in these studies, which would reduce the possibility of finding differences between their results and those of other trials. On the other hand, the probable occurrence of publication bias among trials using inert placebos suggests the effects of antidepressants in these trials might have been overestimated. This study therefore cannot resolve the issue of whether the use of an active placebo reduces the apparent effects of antidepressants. More research using modern trial procedures is required, including investigation of the extent to which 'unblinding' occurs and comparisons of effects using active and inert placebos.

However, the most marked finding from this analysis was the high degree of heterogeneity among results of trials, especially those using ordinary inert placebos. Although some of this was explained by whether trials were conducted with inpatients or outpatients, there was much residual heterogeneity that was not explained by the factors examined. Some of this heterogeneity may derive from the methodological inadequacies of the trials including the inclusion of heterogeneous patient groups and lack of use of standardized diagnostic and outcome instruments. However, as heterogeneity is rarely examined in meta-analyses of more recent trials, it remains a possibility that it exists despite the introduction of more standardized procedures. Heterogeneity among more recent trials needs quantification and exploration. In the meantime, it remains questionable whether any conclusions about the overall effects of antidepressants should be drawn.

Inpatient versus outpatient trials

Contrary to the subsidiary hypothesis of this study, antidepressants were more effective compared with placebos in trials conducted with outpatients. In inpatient trials the difference between antidepressants and any type of placebo was small and not statistically significant. Meta-regression analysis confirmed that whether studies were conducted with inpatients or outpatients strongly predicted outcome. Assuming that patients who require admission to hospital are generally more severely depressed than those who do not, this finding is in contrast with general beliefs that antidepressant-placebo differences are largest in people who are most severely depressed. These beliefs are supported by evidence from individual studies (Paykel et al., 1988; Elkin et al., 1989; Angst et al., 1993) and from two other meta-analyses (Ottevanger, 1993; Khan et al., 2002). Most of this evidence consists of trials conducted with outpatients or patients from general practice. Only one meta-analysis recorded that it included some inpatients (Ottevanger, 1993). In contrast, evidence from a study of inpatients suggested that more severely ill patients responded less well to antidepressants than less severely ill patients (Kocsis et al., 1990).

There are several possible explanations for the apparent inconsistency between the results of this study and most other evidence. Firstly, it is possible that it is patients in the middle of the severity range that respond best to antidepressants compared with placebo. Thus it might be the most severely ill outpatients that respond well to antidepressants and patients with conditions that are severe enough to require admission might respond less well. Secondly, inpatient populations may include patients with psychotic depression, in whom antidepressants are not thought to work well. However, although some studies claimed to include patients with 'psychotic' depression, this label was frequently used to contrast with 'neurotic' depression, and did not necessarily imply the presence of psychotic symptoms. In addition, Kocsis et al. (1990) found that having psychotic symptoms did not predict poor response to antidepressants independently of the severity of depression. Thirdly other factors related to being admitted to hospital may be associated with a poor response to antidepressants, such as co-morbidity or social problems. Fourthly all the inpatient trials in the current study were conducted in the 1960s and it is

difficult to know whether the results in this sample of trials apply to current inpatient populations.

Investigation of heterogeneity

Location of care was the only factor tested that predicted outcome. Overall quality did not predict study result, and nor did individual aspects of quality or year of publication or type and dose of antidepressant used. One previous meta-analysis did find evidence of a relationship between the use of appropriate doses of antidepressants and outcome (Bollini et al., 1999). This analysis also found that not performing intention to treat analysis inflated results. Another meta-analysis, however, failed to find evidence of a dose-response relationship and did not find any association between outcome and aspects of quality, although performing an intention-to-treat analysis was not examined (QAP, 1983). There are therefore few pointers in the existing literature about why results of some antidepressant studies are so variable. In this study, dividing the sample into inpatient and outpatient trials reduced heterogeneity among inpatient trials but it remained high among outpatient trials.

Conclusions

This study did not show a difference between the results of trials of antidepressants using active placebos and trials using ordinary inert placebos. However, it was difficult to resolve the issue of whether unblinding influences outcome because of the limitations of the original trials, the heterogeneity between them and the influence of large studies on combined results. The possible influence of unblinding deserves investigation in more recent trials. An assessment of the integrity of double blind procedures as recommended recently, would be a welcome start (Even et al., 2000).

The most marked finding in this study was high levels of heterogeneity between trials especially trials using inert placebos and outpatient trials, most of which could not be explained. Whether this applies to more recent trials is uncertain, but heterogeneity of antidepressant trials needs further exploration before overall conclusions about the efficacy of antidepressants can be made. Results of other meta-analyses should be appraised with this in mind. Differences between antidepressants and placebo appeared to be greater in outpatients. In this sample of inpatient trials antidepressants seemed to have little effect.

Acknowledgements

I would like to thank Dr Rebecca Hardy for statistical advice and Professor Simon Wessely for looking over an earlier version of this paper.

References

- Abraham HC, Kanter VB, Rosen I, Standen JL. A controlled clinical trial of imipramine with out-patients. *British Journal of Psychiatry* 1963; 109: 286–93.
- Ainslie, JD, Jones MB, and Stiefel JR. Practical drug evaluation method. *Archives of General Psychiatry* 1965; 12: 368–73.
- Anderson IM, Nutt DJ, Deakin JFW on behalf of the Consensus Meeting and endorsed by the British Association of Psychopharmacology. Evidence based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology Guidelines. *Journal of Psychopharmacology* 2000; 14: 3–20.
- Angst J, Scheidegger P, Stabl M. Efficacy of moclobemide in different patient groups: results of new subscales of the Hamilton Rating Scale. *Clinical Neuropharmacology* 1993; 16 (Suppl. 2): S55–S62.
- Bellack AS, Hersen M, Himmelhoch J. Social skills training compared with pharmacotherapy and psychotherapy in the treatment of unipolar depression. *American Journal of Psychiatry* 1981; 138: 1562–7.
- Bollini P, Pampallona S, Tibaldi G, Kupelnick B, Munizza C. Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials. *British Journal of Psychiatry* 1999; 174: 297–303.
- Brick H, Doub WHJ, Perdue WC. Effects of amitriptylene on depressive and anxiety states in penitentiary inmates. *Diseases of the Nervous System* 1962; 22: 572–8.
- Daneman EA. Imipramine in office management of depressive reactions. *Diseases of the Nervous System* 1961; 22: 213–17.
- Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In M Egger, GD Smith, DG Altman (eds) *Systematic Reviews in Health Care*, second edition. London: BMJ Publishing Group, 2001.
- Diamond S. Double-blind controlled study of amitriptylene-perphenazine combination in medical office patients with depression and anxiety. *Psychosomatics* 1966; 7: 371–5.
- Elkin I, Shea T, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB National Institute of Mental Health treatment of depression collaborative research program. *Archives of General Psychiatry* 1989; 46: 971–82.
- Engelhardt DM, Margolis RA, Rudorfer L, Paley HM. Physician bias and the double-blind. *Archives of General Psychiatry* 1969; 20: 315–20.
- Even C, Siobud-Dorocant, Dardennes Critical approach to antidepressant trials. Blindness protection is necessary, feasible and measurable. *British Journal of Psychiatry* 2000; 177: 47–51.
- Fink M, Klein DF, Kramer JC. Clinical efficacy of chlorpromazine-procycidine combination, imipramine and placebo in depressive disorders. *Psychopharmacologia* 1965; 7: 27–36.
- Friedman AS, Granick S, Cohen HW, Cowitz B. Imipramine (tofranil) vs placebo in hospitalized psychotic depressives. *Journal of Psychiatric Research* 1966; 4: 13–36.
- Friedman AS. Interaction of drug therapy with marital therapy in depressive patients. *Archives of General Psychiatry* 1975; 32: 619–37.
- Garry JW, Leonard T J. Trial of amitriptylene in chronic depression. *British Journal of Psychiatry* 1963; 109: 54–5.
- Greenberg RP, Fisher S. Suspended judgement. Seeing through the double masked design: a commentary. *Controlled Clinical Trials* 1994; 15: 244–6.
- Hall J, Lonie D. A controlled trial of amitriptylene for treatment of depression in inpatients. *New Zealand Medical Journal* 1962; 61: 548–50.
- Hamilton MA. A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* 1960; 23: 56–62.
- Hazell P, O'Connell D, Heathcote D, Robertson J, Henry D. Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *BMJ* 1995; 310: 897–901.
- Holdway V. Trial of imipramine. *Journal of Mental Science* 1960; 106: 1443–6.
- 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.
- Hollister LE, Overall JE, Johnson M, Pennington V, Katz G, Shelton J. Controlled comparison of imipramine, amitriptylene and placebo in hospitalised depressed patients. *Journal of Nervous and Mental Disease* 1964; 139: 370–5.
- Hussain Z. Drugs in depressive illness. *British Medical Journal* 1970; 1: 482.
- Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology* 2002; 22: 40–5.
- Kirsch I The emperor's new drugs: an analysis of antidepressant medication data submitted to the US Food and Drug Administration. *Prevention and Treatment*, 2002; 5, available at <http://journals.apa.org/prevention>
- Kocsis JH, Croughan JL, Katz MM, Butler TP, Secunda C, Bowden CL, Davis JM. Response to treatment with antidepressants of patients with severe or moderate non psychotic depression and of patients with psychotic depression. *Am J Psychiatry* 1990; 147: 621–4.
- Kranzler HR, Burleson JA, Korner P, Del Boca FK, Bohn MJ, Brown J, Liebowitz N. Placebo controlled trial of fluoxetine as an adjunct to relapse prevention in alcoholics. *Am J Psychiatry* 1995; 152: 391–7.
- Lorr M, Jenkins R, Holsopple J. Multidimensional scale for

- rating psychiatric patients. Veterans Administration Technical Bulletin. Washington: Veterans Association, 1953; 10–507.
- Master RS. Amitriptyline in depressive states. A controlled trial in India. *British Journal of Psychiatry* 1963; 109: 826–9.
- McCallum P, Reares R. A controlled trial of mapriotiline in depressed out-patients. *Medical Journal of Australia* 1975; 2: 392–3.
- Moncrieff J, Churchill R, Drummond DC, McGuire H. Development of a quality assessment instrument for trials of treatments of depression and neurosis. *International Journal of Methods in Psychiatric Research* 2001b; 10: 126–33.
- Moncrieff J, Wessely S, Hardy R. Meta-analysis of trials comparing antidepressants with active placebos. *British Journal of Psychiatry* 1998; 172: 227–31.
- Moncrieff J, Wessely S, Hardy R. Antidepressants versus active placebos. *Cochrane Database of Systematic Reviews*, 2001a.
- Morris JB, Beck AT. The efficacy of antidepressant drugs. A review of research (1958 to 1972). *Archives of General Psychiatry* 1974; 30: 667–74.
- Murphy GE, Simons AD, Wetzel RD, Lustman PJ. Cognitive therapy and pharmacotherapy. *Archives of General Psychiatry* 1984; 41: 33–41.
- Ottavanger EA. The efficacy of fluvoxamine in patients with severe depression. *British Journal of Clinical Research* 1991; 2: 125–32.
- Overall JE. Dimensions of manifest depression. *Journal of Psychiatric Research*, 1962; 1: 239–248.
- Paykell ES, Hollyman JA, Freeling P, Sedgwick P. Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo controlled trial. *Journal of Affective Disorders* 1988; 14: 83–95.
- Porter AMW. Depressive illness in a general practice. A demographic study and a controlled trial of imipramine. *BMJ* 1970; 1: 773–8.
- Prange AJ, McCurdy RL, Cochrane CM. The systolic blood pressure response of depressed patients to infused norepinephrine. *Journal of Psychiatric Research* 1967; 5: 1–13.
- Quality Assurance Project. A treatment outline for depressive disorders. *Australian and New Zealand Journal of Psychiatry* 1983; 17: 129–46.
- Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF. Validity of clinical trials of antidepressants. *American Journal of Psychiatry* 2000; 157: 327–37.
- Raskin A, Crook TH. Antidepressants in black and white inpatients. *Archives of General Psychiatry* 1975; 32: 643–9.
- Raskin A, Schulterbrandt JG, Reatig N, Chase C, McKeon JJ. Differential response to chlorpromazine, imipramine and placebo. *Archives of General Psychiatry* 1970; 23: 164–73.
- Rickels K, Ward CH, Schut L. Different populations, different responses. A comparative study of two antidepressants each used in two different patient groups. *American Journal of Medical Science* 1964; 47: 328–35.
- Smith RC. Amoxapine, imipramine and placebo in depressive illness. *Current Therapeutic Research* 1975; 18: 346–53.
- Stata 7. Texas, USA: Stata Corporation, 2002.
- Thompson S. Why and how sources of heterogeneity should be investigated. In M Egger, GD Smith, DG Altman (eds) *Systematic Reviews in Healthcare*. London: BMJ Publishing Group, 2001.
- Toneatto T, Sellers EM. Perception of whether drug or placebo has been administered is a determinant of drinking reduction. In EM Sellers and CA Naranjo (eds) *Novel Pharmacological Interventions for Alcoholism*. California: Springer Verlag 1992.
- Uhlenhuth EH, Park LC. The influence of medication (imipramine) and doctor in relieving depressed psychoneurotic outpatients. *Journal of Psychiatric research* 1963; 2: 101–22.
- Weintraub W, Aronson H. clinical judgment in psychopharmacological research. *Journal of Neuropsychiatry* 1963; 5: 65–70.
- 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: 1730–61.
- Wilson IC, Vernon JT, Guin T, Sandifer MG. A controlled study of treatments of depression. *Journal of Neuropsychiatry* 1963; 4: 331–7.

Correspondence: Joanna Moncrieff, Department of Psychiatry and Behavioural Science, University College London, Wolfson Building, Riding House Street, London W1N 8AA.

Telephone (+44) (0) 1277 302695.

Fax (+44) (0) 1277 302696.

Email: j.moncrieff@ucl.ac.uk.