

Randomised controlled trials investigating pharmacological and psychological interventions for treatment-refractory depression

Systematic review

NICOLA STIMPSON, NIRUJ AGRAWAL and GLYN LEWIS

Background About 30% of people with depression do not respond to an antidepressant at the recommended dose and can be described as having treatment-refractory depression.

Aims To summarise the findings from all randomised controlled trials (RCTs) that have assessed the efficacy of a pharmacological or psychological intervention for treatment-refractory depression.

Method We used a systematic search strategy to identify RCTs that included adults aged 18–75 years with a diagnosis of unipolar depression that had not responded to a 4-week course of a recommended dose of an antidepressant.

Results We identified 16 RCTs. None of the included trials assessed the efficacy of psychotherapy. All the trials were too small to detect an important clinical response. We found only two trials on lithium augmentation, which randomised 50 subjects in total.

Conclusions There is little evidence to guide the management of depression that has not responded to a course of antidepressants. Treatment-refractory depression is an important public health problem and large pragmatic trials are needed to inform clinical practice.

Declaration of interest G.L. has received payment for lectures from a number of pharmaceutical companies.

Approximately 30% of people with depressive illness do not respond to the usual recommended dose of antidepressants. The World Psychiatric Association made one of the earliest definitions of ‘resistant’ depression as, ‘an absence of clinical response to treatment with a tricyclic antidepressant at a minimum dose of 150 mg/day of imipramine (or equivalent drug) for 4 to 6 weeks’ (World Psychiatric Association, 1974). A number of alternative definitions have been used but the term ‘treatment-refractory depression’ that we adopt here will be the World Psychiatric Association definition with a 4-week time criterion. Most other definitions require more ‘severe’ treatment-refractory depression, in the sense that patients have failed to respond to more than a single course of antidepressant (Thase & Rush, 1995).

Current guidance

There is little current guidance on the management of treatment-refractory depression. Current guidelines (American Psychiatric Association, 1993; Anderson *et al*, 2000) suggest increasing the dose of antidepressant, switching to a different class, adding psychotherapy or augmenting with lithium or electroconvulsive treatment. The lack of guidance is reflected by variation in the management of treatment-refractory depression. A third of psychiatrists in the north-east of the USA preferred lithium augmentation (Nierenberg & White, 1990). Canadian psychiatrists (Chaimowitz *et al*, 1991) had an equal preference for a second tricyclic, augmentation with a monoamine oxidase inhibitor and augmentation with lithium. The most popular choice in the UK (Shergill & Katona, 1996) was to increase the dose or to change class. However, 39% of respondents in this study stated that they were not confident when treating this condition.

Previous systematic reviews

Systematic reviews of the literature attempt to provide an unbiased and succinct summary of all of the available evidence and, when possible, produce a meta-analysis that summarises results more precisely (Chalmers & Altman, 1995; Lewis *et al*, 1997). Previous systematic reviews have assessed the efficacy of lithium augmentation (Austin *et al*, 1991; Bauer & Dopfmer, 1999) and triiodothyronine augmentation (Aronson *et al*, 1996). The systematic review of Austin *et al* included 5 trials, but 4 of these used only 3 weeks to define treatment resistance. One of the trials treated subjects with lithium for only 48 hours, and another reported very low (less than 0.3 mmol/l) blood lithium levels. Bauer & Dopfmer (1999) included randomised controlled trials (RCTs) in their review that studied both unipolar and bipolar depression. It would seem unwise to generalise from patients with bipolar depression to those with unipolar depression, especially in relation to lithium use. The systematic review of four randomised double-blind studies of triiodothyronine (Aronson *et al*, 1996) also included studies that used a 3-week criterion and patients with bipolar depression.

The aim of this systematic review was to identify and summarise all the RCTs that had investigated the pharmacological and psychological management of patients with treatment-refractory depression.

METHOD

A literature search was carried out in association with the Cochrane Collaboration (Depression, Anxiety and Neurosis Group). The Cochrane Controlled Trials register (CTTR) 2000 edition was searched, as were the following electronic databases: EMBASE (1980–1999), Medline (1966–1999), Psychlit and PsychInfo (1974–1999), LILACS (1982–1999). The standard search strategy for identifying RCTs developed by the Cochrane Collaboration was used (<http://www.cochrane.org>). Keywords to identify treatment-refractory depression trials include DEPRESS*; THERAPY or TREATMENT, REFRACT*; RESISTANT; NON-RESPOND*; UN-RESPONS*; FAIL*; AUGMENT*; POTENTIATION and COMBIN*. The abstracts of these trials were read to identify those that appeared to reach the inclusion criteria. Paper or electronic copies

of trials that appeared, from the abstract, to achieve the inclusion criteria were collected for further inspection.

When the search strategy had been completed, the authors of all identified trials (both those to be included and the 'near misses') and all known experts in the field were contacted for any further information on trials that were unpublished, in press or were currently in progress. If trials presented data on both unipolar and bipolar depression the authors were asked for the results of the unipolar participants.

Inclusion criteria

Randomised controlled trials were included in the review if the participants had a diagnosis of unipolar depression that had not responded to a minimum of 4 weeks of antidepressant treatment at a recommended dose (at least 150 mg/day imipramine or equivalent). This definition was chosen in order to include as much evidence as possible. Trials that concentrated solely on patient groups either under the age of 18 years or over the age of 75 years were excluded, as were trials including patients with comorbid schizophrenia. Participants with bipolar disorder were excluded. These criteria and the details of the search strategy were decided before beginning the review and published as a protocol in the Cochrane Database of Systematic Reviews (Stimpson *et al*, 2000).

Summary data from each of the identified trials were extracted independently by at least two of the three reviewers and entered onto predesigned data extraction forms. Any disagreements were discussed until a consensus was reached. If additional information was needed the first author of the trials was contacted.

Statistical methods

Where possible we planned to carry out meta-analysis of the results from trials. We wished to use a dichotomous outcome, the numbers who had 'recovered'. This is usually reported as a 50% reduction in Hamilton Rating Scale for Depression (HRSD) scores (Hamilton, 1960). This outcome was chosen for two main reasons. First, it avoids the difficulty of establishing whether a continuous variable has a normal distribution. Second, it allows fairly simple analyses that aid interpretation, particularly from a clinical perspective. We chose to calculate the absolute risk difference (i.e. the difference in proportion recovered).

The reciprocal of this measure is the number needed to treat (Sackett & Cook, 1995). A positive value for a risk difference was given when the proportion recovered was greater in the intervention than in the placebo group. For the small trials, exact confidence intervals were calculated. Otherwise, risk difference, 95% confidence intervals and tests for heterogeneity were calculated using the Metan command within Stata (StataCorp, 1999).

RESULTS

Using our search strategy, 753 potential trials were initially identified and this number increased as the search was updated quarterly until January 2001 to give a total

of 919 trials. Forty studies were excluded from the review, in accordance with our published protocol (Stimpson *et al*, 2000). The search and identification of studies is summarised in Fig. 1.

Exclusions

Fourteen trials were excluded from the review as they included participants with unipolar and with bipolar depression and it was not possible to extract data on unipolar depression alone. In 11, participants had been on antidepressant medication for less than 4 weeks or at a dose of less than 150 mg imipramine or equivalent. Three trials were abandoned on the grounds of the randomisation. In one relevant trial

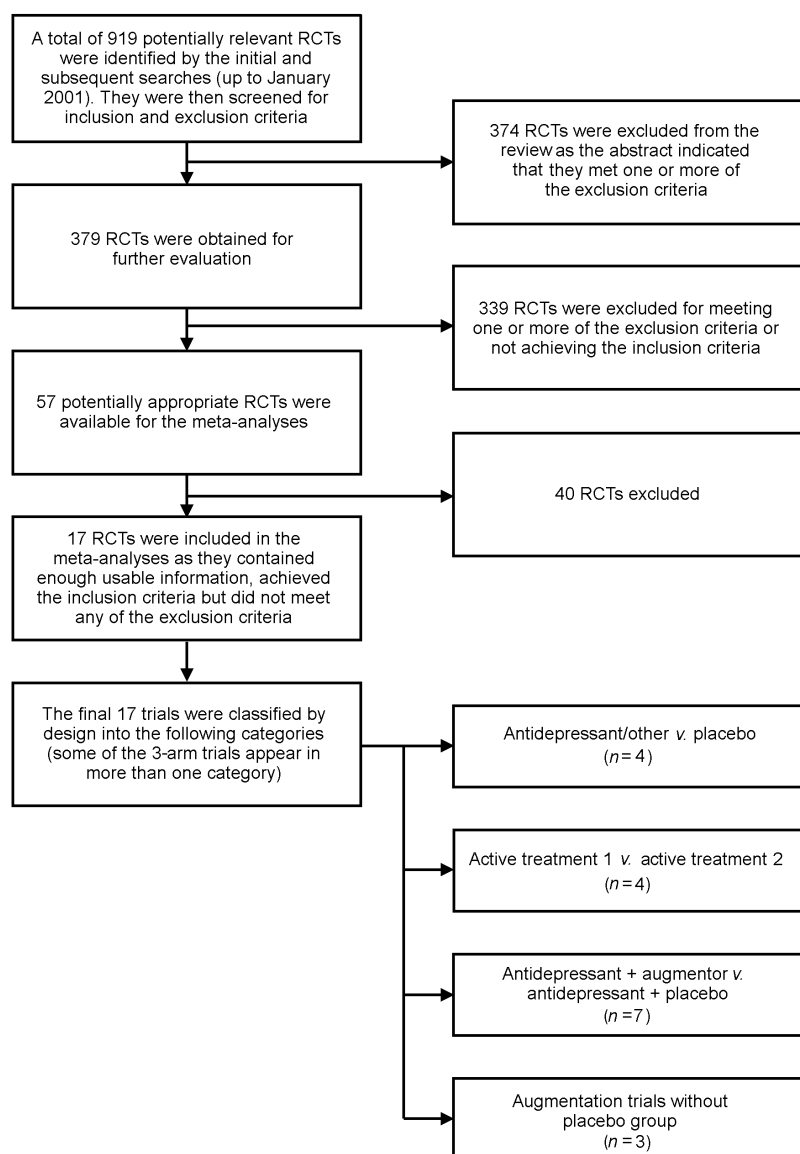


Fig. 1 Flowchart of progress through systematic review. RCTs, randomised controlled trials.

the randomisation had given rise to a striking imbalance between the randomised groups (Gitlin *et al*, 1987). This may well have resulted from the small size of these trials ($n=16$). One trial randomised participants to identical treatments (Antonuccio *et al*, 1984). A full list of excluded studies is available from the author upon request.

Two crossover trials were also excluded because it was impossible to extract data from the initial phase of the trial before the crossover took place. One published (Gagiano *et al*, 1993) and one unpublished trial (source available from the author upon request) had to be excluded as they did not describe the study with sufficient detail to know whether the inclusion criteria were met. One trial had to be excluded as data were not available on the subset of participants that were randomly assigned to cognitive-behavioural therapy (Barker *et al*, 1987). Two papers presented previously published results and the duplicated results are not included in the review (Zohar *et al*, 1985; Joffe & Singer, 1992).

Included trials

Seventeen RCTs were identified, which included a total of 645 participants. A variety of different designs were adopted. After extracting the data we have chosen to classify these designs according to the following four categories.

Antidepressant (or other) v. placebo (Table 1)

There were four trials which compared a pharmacological agent with a placebo (Table 1). The agents investigated were oestrogen (Klaiber *et al*, 1979), viqualine (Faravelli *et al*, 1988), ketoconazole (Malison *et al*, 1999) and paroxetine (Tyrer *et al*, 1987). Two of these studies were also crossover trials from which we extracted data for the 2 weeks prior to crossover.

Two of these trials (Klaiber *et al*, 1979; Faravelli *et al*, 1988) found a significant advantage compared with placebo, despite their low statistical power. The largest of these four trials randomised 47 subjects. In three trials that reported recovery rates, none of the 38 subjects randomised to placebo recovered (97.5% CI 0–9%).

We excluded the results from the second phase of the crossover designs.

Comparison of two active treatments

There were four trials that compared two pharmacological agents (Table 2). The comparisons made were: intravenous maprotiline v. intravenous clomipramine (Drago *et al*, 1983); brofaromine v. tranylcypromine (Nolen *et al*, 1993); venlafaxine v. paroxetine (Poirier & Boyer, 1999); and olanzapine v. fluoxetine (Shelton *et al*, 2001).

The venlafaxine v. paroxetine comparison seems most relevant to current practice. The results of this trial did not support the superiority of one or other compound. Three of the performed analyses led to a result that favoured venlafaxine, but two of these did not adopt an intention-to-treat policy and most were of marginal statistical significance. Almost two-thirds of the subjects had been on a selective serotonin reuptake inhibitor previously. The Shelton study examined the policy of 'switching' between fluoxetine and olanzapine as all the subjects had failed to respond to fluoxetine. There was little information on previous medication for the other studies.

Antidepressant+augmenter v. antidepressant+placebo

The comparison of an augmentation strategy with a placebo seems the most relevant to clinical practice. Two trials of lithium as an augmentation agent (Zusky *et al*, 1988; Joffe *et al*, 1993; Table 3) could be included and a meta-analysis performed. In summary, lithium had a recovery rate by the end of the trial 25% greater than placebo (95% CI 2–49%), corresponding to a number needed to treat of 4 (95% CI 2–50). In all, there were only 50 patients in the two lithium trials. There was no statistical evidence to support heterogeneity between the trials ($\chi^2=0.6$, d.f.=1, $P=0.44$).

There were also three trials of pindolol as an augmenter (Maes *et al*, 1996; Moreno *et al*, 1997; Perez *et al*, 1999) reporting on 106 subjects, although one of these (Moreno *et al*, 1997) did not report any recoveries and therefore does not contribute towards the summary estimate. Overall, those given pindolol had an 8% better recovery rate (95% CI 21% to –6%) but this was not statistically significant. There was little evidence to support any heterogeneity between the three pindolol trials ($\chi^2=5.46$, d.f.=2, $P=0.07$). Three further trials also used this design but investigated different augmentation strategies

(Maes *et al*, 1996; Clifford *et al*, 1999; Shelton *et al*, 2001).

The overall recovery rate on placebo in all the eight trials was 14 out of 107 subjects or 14.4% (95% CI 7.9–23.4%).

Augmentation without a placebo

There were three trials that investigated augmentation of an antidepressant but did not compare with a placebo (Joffe & Singer, 1990; Fava *et al*, 1994; Rybakowski *et al*, 1999) (Table 4).

Methodological quality of trials

None of the trials would have met all the requirements of the CONSORT guidelines on reporting results of randomised trials (Begg *et al*, 1996). Two of the trials mentioned that the random numbers were generated with a computer program. Of the ten trials that used a placebo, four mentioned that the placebos were identical in appearance to the active treatment. None of the trials gave an indication of how the allocation of randomisation was conducted, and only one trial (Perez *et al*, 1999) described how the randomisation was concealed. The two lithium trials mentioned that faked blood results were used to maintain blindness.

Four studies (Joffe & Singer, 1990; Joffe *et al*, 1993; Perez *et al*, 1999; Poirier & Boyer, 1999) reported a power calculation, although one reported a power of 20%. Two trials recruited the exact number of participants required by their power calculations (Joffe & Singer, 1990; Perez *et al*, 1999). One trial reported that the small sample size recruited had limited the power of their trial (Joffe *et al*, 1993) and one trial reported a power calculation incorrectly and did not report the sample size it required (Poirier & Boyer, 1999). The size of the randomised groups ranged from a maximum of 62 participants to a minimum of 5 participants. Only 2 of the 17 trials had a group with 25 or more subjects.

Issues not addressed by studies

No RCTs were identified that assessed the efficacy of psychotherapy and also met the inclusion criteria. A number of trials of psychotherapy were excluded on various grounds (further details available from the author upon request).

No RCTs were identified that investigated increasing the dose of antidepressant, or that compared switching to a new class

Table 1 Antidepressant/other v. placebo

Study	Participants randomised (n) and method of randomisation	Participants	Definition of 'refractory' (or equivalent term)	Interventions	Duration of intervention	Recovered		Absolute risk difference (95% CI)	Comments
						Group 1	Group 2		
Klaiber <i>et al</i> , 1979	47 'double-blind randomised assignment'	Female in-patients with severe depression. Score of at least 25 on HRSD and maintenance of at least 20 on HRSD for the placebo period	Failed to respond to various conventional treatments of depression for at least 2 years	Placebo period of 5–6 weeks. Group 1, oestrogen 5–25 mg/day or, Group 2, placebo (of similar appearance). Previous medications were withdrawn	12 weeks	11/25	0/22	44% (25 to 63%)	Improvement defined as 10 or more points reduction on HRSD
Faravelli <i>et al</i> , 1988	20 'double-blind placebo-controlled trial assigned on a random basis'	In-patients with a diagnosis of major depressive episode (DSM-III-R). Score of at least 18 on HRSD, age between 18 and 65 years	Previous episodes of major depression in which the patients showed poor response to treatment with tricyclic antidepressants at doses of at least 150 mg	Group 1, lorazepam (2.5 mg 4/day) plus viqualine (50 mg/day), or Group 2, lorazepam (2.5 mg 4/day) plus placebo (4 capsules/day)	4 weeks	5/10	0/10	50% (19 to 81%)	Some participants had dysthymia
Malison <i>et al</i> , 1999	16 'randomly assigned under double-blind conditions'	Participants met diagnostic criteria for major depressive disorder (DSM-III-R) and had a score of at least 17 on a modified 19-item HRSD	Non-response to two different antidepressants or to one antidepressant with lithium augmentation. The average participant greatly exceeded these criteria	Group 1, ketoconazole (titrated from 200 mg/day to a maximum of 1200 mg/day) or, Group 2, placebo (identical appearing capsules and doses)	6 weeks	1/7	0/6	14% (–12 to 40%)	Results shown include drop-outs but exclude with bipolar disorder
Tyrer <i>et al</i> , 1987	37 'double-blind procedure'	Out-patients with a primary depressive illness (RDC). Had a score of at least 15 on 21-item HRSD. Had at least one other episode of depressive illness lasting for 3 months or longer, or if this episode was the first, had treatment for 3 months or longer without response	Had received a tricyclic antidepressant for at least 4 weeks in therapeutic dosage and failed to show a clinical response or had to stop treatment because of unwanted effects	2-week placebo phase. Group 1, paroxetine (30 mg/day), or Group 2, placebo (of identical appearance)	2 weeks	Baseline 20.0 (0.9); 4 weeks 15.4 (1.8) ¹	Baseline 21.1 (1.4); 4 weeks 19.6 (1.4) ¹		Crossover design with data extracted from period before crossover (4 weeks)

HRSD, Hamilton Rating Scale for Depression; RDC, research diagnostic criteria.

1. Results given as continuous measure: group HRSD score at baseline (s.e.) and at end of 4 weeks (s.e.).

Table 2 Antidepressant v. antidepressant 2

Study	Participants randomised (n); method of randomisation	Participants	Definition of 'refractory' (or equivalent term)	Interventions (or equivalent term)	Duration of intervention	Recovered		Absolute risk difference (95% CI)	Comments
						Group 1	Group 2		
Drago <i>et al.</i> , 1983	40; 'double-blind randomised trial'	Female in-patients aged between 18 and 70, diagnosed with resistant and severe primary depression. Score of at least 18 on 17-item HRSD, or overt suicidal tendency	Patient had not responded to 3 different major antidepressant agents in the previous 6 months	Group 1, intravenous maprotiline (100 mg/day), or Group 2, intravenous clomipramine (100 mg/day)	3 weeks	10/20	14/20	-20% (-50 to 10%)	Data extracted up to the ninth day of the trial as participants were crossed over at that point
Nolen <i>et al.</i> , 1993	36; 'randomly assigned to a double-blind treatment'	Participants met DSM-III-R criteria for major depression and scored at least 18 on the HRSD (17-item)	Less than a 50% reduction on HRSD baseline score after 4 weeks of treatment with maprotiline or nortriptyline	Group 1, brofaromine (50 mg), or Group 2, tranylcypromine (20 mg)	4 weeks	9/20	5/16	14% (-18 to 45%)	Extra data provided by author for the participants with unipolar depression. Three participants with bipolar depression excluded from analysis
Poirier & Boyer, 1999	123; 'double-blind randomised multicentre comparison'. Assigned a random number and randomised in blocks of four at each treatment centre	In-patients or out-patients aged 18-60 who satisfied DSM-III-R criteria for major depression and whose depression was less than 8 months old. Score of at least 18 on 17-item HRSD	Participants required to have a history of resistance to two previous successive antidepressant treatments for the current episode. The first treatment had to have been for at least 4 weeks at an effective dose	Group 1, venlafaxine (started at 37.5 mg twice daily), up to a maximum of 200-300 mg/day, or Group 2, paroxetine (20 mg/day) up to a maximum of 30-40 mg/day	4 weeks	27/61*	18/62*	15% (-2 to 32%)	Last observation carried forward analysis used. Participants also had to be 'very much improved' on the Clinical Global Impression Scale
Shelton, 2001	18; 'randomly assigned' to the 'double-blind trial'	Out-patients who achieved DSM-IV criteria for depression; 75% were women and 96% were White. Mean age was 42 years (s.d.=11). Score of ≥ 20 on 21-item HRSD	Treatment resistance defined as a 'history of failure to respond to antidepressants of two different classes, one of which was not an SSRI, after at least 4 weeks of therapy at an acceptable therapeutic dose'	Group 1, olanzapine (5-20 mg/day) plus placebo or, Group 2 fluoxetine (20-60 mg/day) plus placebo	8 weeks	0/8	1/10	-10% (-29 to 9%)	

HAM-D, Hamilton Rating Scale for Depression; SSRI, selective serotonin reuptake inhibitor.

* $p < 0.05$.

of antidepressant with remaining on the original antidepressant.

DISCUSSION

Only 17 RCTs were identified, including 645 participants, covering any pharmacological or psychological intervention for treatment-refractory depression. The most striking impression is that there is currently very little evidence to guide the management of those who have not responded to a standard dose of antidepressant for 4 weeks. Augmentation of existing antidepressant medication was the strategy that had received most investigation, whereas there were no studies of any psychological treatment. It was possible to conduct a meta-analysis with the results from two trials that investigated lithium and the three that studied pindolol. The remaining studies mostly investigated a range of therapeutic options that, overall, did not address questions of current clinical relevance. Treatment-refractory depression is a common clinical problem and this lack of evidence is reflected in an absence of consensus among clinicians and the vagueness of current guidelines.

Methodology

The systematic review used a thorough search strategy as part of the Cochrane Collaboration. It is still possible, however, that some trials have not been identified despite our efforts, and we would welcome any information about trials, particularly those that are unpublished.

The major limitation of the review reflects the major weakness of the constituent trials. Almost all the studies were small in size. Only 2 of the 17 trials had 25 or more subjects in a randomised group. A trial with 25 subjects in each group would be able to detect the difference between 10% and 50% recovery with 80% power and 5% significance. This is a large difference in outcome, much larger than the 14% difference reported in a recent meta-analysis of fluoxetine *v.* placebo (Bech *et al.*, 2000). A trial would have to randomise 219 subjects to each group to detect a difference between 10% and 20% recovery with 80% power and 5% significance. All the trials in this study were therefore severely underpowered. Small trials can also lead to a failure of randomisation, resulting in an imbalance between the randomised groups. We came across two studies where

this had occurred and excluded them, but smaller degrees of imbalance might still be present.

Publication bias was impossible to assess as the trials studied such a diverse range of interventions. It is usually assumed that systematic reviews of small trials are likely to be more susceptible to publication bias than those that include larger trials. Even meta-analysis of moderately sized trials can provide biased conclusions (LeLorier *et al.*, 1997).

Since 1996, the CONSORT statement has provided guidance on the reporting of RCTs (Begg *et al.*, 1996). None of the 17 studies, including those published after the CONSORT statement, followed all aspects of its guidance. Trials with inadequate concealment of allocation are associated with an increased estimate of benefit (Moher *et al.*, 1998). Only one trial described how they kept the allocation of subjects concealed from the clinicians involved in their care (Perez *et al.*, 1999). Overall, the trials did not meet the current expectations concerning the adequate reporting of randomised trials.

Inclusion criteria

The World Psychiatric Association (1974) defined treatment-refractory depression as a failure to respond after a 4- to 6-week period on a recommended dose of antidepressant. When planning the review and without prior knowledge of the included studies, we chose to set our inclusion criteria using a time limit of 4 weeks. This minimum time limit was considered appropriate for a systematic review as it would ensure that we collected all relevant studies. It also reflected the commonest clinical dilemma: what to do next after lack of response to an antidepressant. We were surprised that we excluded nine trials on the grounds that they defined treatment-refractory depression using a time limit of 3 weeks. Because the response to antidepressants can be delayed, we think this definition is rather too broad. We also excluded 14 trials on the grounds that they included both patients with bipolar and with unipolar depression. The management of depression in those with bipolar depression differs in some important respects from those with unipolar depression. Antidepressants are used more cautiously in case this precipitates a manic relapse. In the context of a trial, a manic relapse might lead to an apparent 'improvement' in depression

scores. Most people with established bipolar disorder would also be on a mood stabiliser such as lithium.

Design of trials

We excluded the second phase of crossover designs as these are inappropriate for antidepressant trials in which subjects may recover. Antidepressants have a delay of 2–3 weeks before they take effect and so short periods before crossover are uninformative, as acknowledged by Tyrer *et al.* (1987).

We identified four different designs in our included studies. Four studies compared an antidepressant *v.* a placebo, thus investigating removing an antidepressant agent and replacing with placebo. Because some subjects with 'treatment-refractory depression' will have had a partial response, removal of antidepressant would be expected to lead to a worsening of symptoms. Two of the four trials using this design found improved recovery on active antidepressant. These results argue against stopping antidepressant medication in those who have not had a good response.

Four trials compared two active treatments. This also investigates switching to another antidepressant following failure to respond. However, the most relevant trial (Poirier & Boyer, 1999), which compared venlafaxine and paroxetine, included subjects that had been exposed to either selective serotonin reuptake inhibitors, tricyclics or both. To study the policy of switching to a new antidepressant, a more informative design would be to recruit subjects who had been treated with a single class of antidepressant and then randomise to either staying on the same class of antidepressant or switching to an alternative class. This design was used (Shelton *et al.*, 2001) to compare remaining on fluoxetine with switching to olanzapine.

Augmentation

The most informative designs were those in which an augmenting agent was added to antidepressant medication and compared with a placebo and antidepressant. Our finding that 14% (95% CI 8–23%) of the placebo group recovered emphasises the necessity of a placebo comparison for studies of augmentation.

The two lithium trials were small, with only 50 patients in all, and treated subjects for 1–2 weeks, a relatively short duration. Although there was a statistically

Table 3 Antidepressant+augmentor v. antidepressant+placebo

Study	Participants randomised (n); method of randomisation	Participants	Definition of 'refractory' (or equivalent term)	Interventions	Duration of intervention	Recovered		Absolute risk difference (95% CI)	Absolute risk difference of meta-analysis (95% CI fixed)	Comments
						Group 1	Group 2			
Zusky <i>et al.</i> , 1988	16; 'double-blind placebo-controlled randomised trial'	Participants met DSM-III-R criteria for major depression without psychosis. Aged 18–80 years. HRSD score of 12 or higher	Failed to respond to a minimum of 4 weeks of treatment with 150 mg or greater of imipramine or its equivalent	Group 1, lithium (300–900 mg/day), 0.1–0.8 mmol/l, or Group 2, placebo (same number of tablets). Participants remained on a range of antidepressants throughout the trial	2 weeks	2/8	1/8	12% (–25 to 50%)		
Joffe <i>et al.</i> , 1993	34; 'double-blind placebo-controlled randomised trial'	Participants met RDC criteria for major, non-psychotic unipolar major depression. Scored at least 16 on the HRSD (17-item)	Failed to respond to a 5-week trial of desipramine or imipramine	Group 1, lithium (900–1200 mg/day), 0.56–0.93 mmol/l, or Group 2, placebo (same number of unmarked capsules). Participants remained on desipramine or imipramine throughout the trial	2 weeks	9/18	3/16	31% (1 to 61%)		
Results of the two lithium trials (meta-analysis)	50					11/26	4/24		25% (2 to 49%)	
Moreno <i>et al.</i> , 1997	10; 'randomised double-blind placebo-controlled crossover trial'	Out-patients with DSM-III-R diagnosis of major depressive disorder and score of at least 18 on 25-item HRSD	Failed to obtain or maintain a therapeutic response to at least 8 weeks of antidepressant medication	Participants were randomised to either Group 1, pindolol (2.5 mg 3/day), or Group 2, placebo. Both groups remained on their current antidepressant	2 weeks	0/5	0/5	0%		Crossover design. Data extracted at 2 weeks (before crossover)
Perez <i>et al.</i> , 1999	80; 'double-blind randomised placebo-controlled trial using computer-generated random digits'	Existence of a major depressive disorder (DSM-IV) and current episode resistant. Score of > 16 on 17-item HRSD	Minimum of 6 weeks' pharmacological treatment of specified dosage	After a 5-day placebo phase, participants were randomised to either: Group 1, pindolol (2.5 mg 3 times a day), or Group 2, placebo (identical tablets 3 times a day). Both groups maintained existing antidepressant medication	10 days	5/40	5/40	0% (–14 to 14%)		

Table 3 (continued)

Study	Participants randomised (n); method of randomisation	Participants	Definition of 'refractory' (or equivalent term)	Interventions	Duration of intervention	Recovered		Absolute risk difference (95% CI)	Absolute risk difference of meta-analysis (95% CI fixed)	Comments
						Group 1	Group 2			
Maes <i>et al</i> , 1996	16; 'randomised using a double-blind placebo-controlled design'	Participants met DSM-III-R diagnostic criteria for major depression. Aged between 25 and 70 years	A minimum of two adequate trials with antidepressant agents from different classes	After a 10 day wash-out period of all antidepressant medication, participants were randomised to either: Group 1, trazodone (100 mg) plus pindolol (7.5 mg/day) or, Group 2, trazodone (100 mg) plus placebo	4 weeks	5/8	1/8	50% (9 to 90%)		Only data on completers could be extracted
Results of the 3 pindolol trials (meta-analysis)	106					10/53	6/53	8% (-6 to 21%)		
Maes <i>et al</i> , 1996	18; 'randomised using a double-blind placebo-controlled design'	Participants met DSM-III-R diagnostic criteria for major depression. Aged between 25 and 70 years	A minimum of two adequate trials with antidepressant agents from different classes	10-day wash-out period of all antidepressant medication. Group 1, trazodone (100 mg) plus fluoxetine (20 mg/day) or, Group 2, trazodone (100 mg) plus placebo	4 weeks	7/10	1/8	58% (21 to 94%)		Numbers reflect only completers
Clifford <i>et al</i> , 1999	10; 'randomly allocated ... to a double-blind placebo-controlled trial'	Participants met DSM-IV criteria for major depression. Mean age 41 years (range 29-56)	Had not made a 'clinically satisfactory response' to therapeutic doses of serotonergic antidepressants	Group 1, buspirone plus penbutolol (40 mg/day) or, Group 2, buspirone plus placebo	10 days	2/5	2/5	0% (-61 to 61%)		
Shelton, 2001	20; 'randomly assigned' to the 'double-blind trial'	Out-patients met DSM-IV criteria for depression. 75% were women and 96% were White. Mean age 42 years (s.d.=11). Score ≥ 20 on 21-item HRSD	Treatment resistance defined as a 'history of failure to respond to antidepressants of 2 different classes, one of which was not an SSRI, after at least 4 weeks of therapy at an acceptable therapeutic dose'	Group 1, fluoxetine (20-60 mg/day) plus olanzapine (5-20 mg/day) or, Group 2, fluoxetine (20-60 mg/day), plus placebo	8 weeks	6/10	1/10	50% (14 to 86%)		

HRSD, Hamilton Rating Scale for Depression; RDC, research diagnostic criteria; SSRI, selective serotonin reuptake inhibitor.

Table 4 Augmentation trials without placebo group

Study	Participants randomised (n); method of randomisation	Participants	Definition of 'refractory' (or equivalent term)	Interventions	Duration of intervention	Recovered			Absolute risk difference (95% CI)	Comments
						Group 1	Group 2	Group 3		
Fava <i>et al</i> , 1994	41; 'double-blind controlled study, participants were randomly assigned'	Participants met criteria for major depressive disorder (DSM-IV-R), score of 16+ on 17-item HRSD, aged 18–65 years	Failed to achieve a 50% or greater reduction in HRSD score with 8 weeks of treatment with 20 mg/day of fluoxetine	Group 1, high dose fluoxetine (40–60 mg/day) or, Group 2, fluoxetine (20 mg/day) plus desipramine (25–50 mg/day) or, Group 3, fluoxetine (20 mg/day) plus lithium (300–600 mg/day)	4 weeks	8/15	3/12	4/14	–4% (–38 to 30%)	Group 2 and 3 compared as fluoxetine dose was identical for both groups
Rybakowski, 1999	41; 'randomly allocated'	Participants met ICD-10 and DSM-IV criteria for depression. Aged 20–70 years. Score of 18 or more on 17-item HRSD	Participants had to show a 'failure to respond to two adequate courses of antidepressant treatment for current episode'	Group 1, lithium (mean daily dose of 965 mg, range 500–1500 mg); dose then adjusted to maintain plasma lithium concentration in the range of 0.5–0.8 mEq/l, or Group 2, carbamazepine (400 mg/day) adjusted to maintain plasma carbamazepine concentration of 4–8 µg/ml. Remained on current antidepressant medication	4 weeks	14/21	13/20	N/A	1.6% (–27 to 31%)	Additional information on unipolar participants provided by authors. Data on 18 bipolar participants excluded from analysis
Joffe & Singer, 1990	40; 'double-blind trial, randomisation by computer-generated random digits'	Attendees at a clinic for primary, non-psychotic, major depressive disorder (RDC). Score of at least 16 on 17-item HRSD	Failed an adequate trial of desipramine or imipramine (min. 4 weeks at dose of 2.5–3 mg/kg body weight (approx.)	Group 1, T ₃ (37.5 µg/day) or Group 2, T ₄ (150 µg/day); remained on current antidepressant medication	3 weeks	7/19	4/21	N/A	18% (–10 to 45%)	Comparison of T ₃ or T ₄ as augmenting agents

HRSD, Hamilton Rating Scale for Depression; RCD, research diagnostic criteria; T₃, triiodothyronine; T₄, thyroxine.
 1. Group 2 v. group 3.

significant benefit for lithium, the confidence intervals are so wide (2–49%) that it does not exclude an inconsequential benefit. Meta-analysis of small trials often leads to unreliable results as randomisation is less effective and publication bias more common. These studies provide very weak evidence to support the use of lithium, although it is a common strategy and has widespread clinical support.

Pindolol is a β -adrenoceptor/5-HT_{1A} receptor antagonist and has been investigated as an augmentation agent in three randomised trials. Overall, there was no significant benefit demonstrated in these three trials. In aggregate, only 106 patients were studied and the wide confidence intervals did not exclude the possibility that pindolol would be an effective augmenting agent.

Further research

The results of our review support the view that further RCTs need to be conducted to investigate the management of treatment-refractory depression. The STAR*D project (<http://www.edc.gsph.pitt.edu/stard/>) funded by the US National Institute of Mental Health will hopefully address a number of the deficiencies in the current literature. We suggest that future RCTs should concentrate on studying the effectiveness of psychotherapy as it is a popular and acceptable option for many patients. The second area of research should be into augmentation strategies. Lithium is supported by the most encouraging results at present, but the evidence is still weak. Further trials should estimate the likely benefits of lithium more accurately and also attempt to refine the indications for its use.

REFERENCES

American Psychiatric Association (1993) Practice guideline for major depressive disorder in adults. *American Journal of Psychiatry*, **150**, 1–26.

Anderson, I. M., Nutt, D. J. & Deakin, J. F. (2000) Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines. British Association for Psychopharmacology. *Journal of Psychopharmacology*, **14**, 3–20.

Antonuccio, D. O., Akins, W. T., Chatham, P. M., et al (1984) An exploratory study: the psychoeducational group treatment of drug-refractory unipolar depression. *Journal of Behavioural Therapy and Experimental Psychiatry*, **15**, 309–313.

Aronson, R., Offman, H. J., Joffe, R. T., et al (1996) Triiodothyronine augmentation in the treatment of

CLINICAL IMPLICATIONS

- Treatment-refractory depression is common in clinical practice but there is little evidence to inform management.
- There was some evidence of benefit for lithium augmentation, but the evidence was very weak.
- In the absence of good evidence, clinicians will have to rely upon their own clinical judgement in deciding upon treatment.

LIMITATIONS

- Like all systematic reviews it is limited by the quality of the constituent studies.
- The main conclusion is that further research is required as the findings are not strong enough to support any clinical guidance.
- It proved difficult to perform much quantitative synthesis because the interventions were so diverse.

NICOLA STIMPSON, BSc, University of Wales College of Medicine, Department of Psychological Medicine, Cardiff; NIRUJ AGRAWAL, MRCPsych, GLYN LEWIS, FRCPsych, Division of Psychiatry, University of Bristol, Bristol, UK

Correspondence: Professor Glyn Lewis, Division of Psychiatry, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, UK. E-mail: glyn.lewis@bristol.ac.uk

(First received 29 November 2001, final revision 14 May 2002, accepted 17 May 2002)

refractory depression. A meta-analysis. *Archives of General Psychiatry*, **53**, 842–848.

Austin, M. P., Souza, F. G. & Goodwin, G. M. (1991) Lithium augmentation in antidepressant-resistant patients. A quantitative analysis. *British Journal of Psychiatry*, **159**, 510–514.

Barker, W. A., Scott, J. & Eccleston, D. (1987) The Newcastle chronic depression study: results of a treatment regime. *International Clinical Psychopharmacology*, **2**, 261–271.

Bauer, M. & Dopfmer, S. (1999) Lithium augmentation in treatment-resistant depression: meta-analysis of placebo-controlled studies. *Journal of Clinical Psychopharmacology*, **19**, 427–434.

Bech, P., Cialdella, P., Haugh, M. C., et al (2000) Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *British Journal of Psychiatry*, **176**, 421–428.

Begg, C., Cho, M., Eastwood, S., et al (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, **276**, 637–639.

Chaimowitz, G. A., Links, P. S., Padgett, R. W., et al (1991) Treatment-resistant depression: a survey of practice habits of Canadian psychiatrists. *Canadian Journal of Psychiatry*, **36**, 353–356.

Chalmers, I. & Altman, D. G. (1995) *Systematic Reviews*. London: BMJ.

Clifford, E. M., Whale, R., Sharp, T., et al (1999) Effect of penbutolol on neuroendocrine responses to buspirone and mood in SSRI-refractory patients: A preliminary study. *Human Psychopharmacology*, **14**, 273–277.

Drago, F., Motta, A. & Grossi, E. (1983) Intravenous maprotiline in severe and resistant primary depression: a double-blind comparison with clomipramine. *Journal of International Medical Research*, **11**, 78–84.

Faravelli, C., Albanesi, G. & Sessarego, A. (1988) Viqualine in resistant depression: a double-blind, placebo-controlled trial. *Neuropsychobiology*, **20**, 78–81.

Fava, M., Rosenbaum, J. F., McGrath, P. J., et al (1994) Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: a double-blind, controlled study. *American Journal of Psychiatry*, **151**, 1372–1374.

Gagliano, C. A., Muller, P. G. M., Gourie, J., et al (1993) The therapeutic efficacy of paroxetine: (a) an open study in patients with major depression not responding to antidepressants; (b) a double blind comparison with amitriptyline in depressed outpatients. *Acta Psychiatrica Scandinavica*, **80**, 130–131.

Gitlin, M. J., Weiner, H., Fairbanks, L., et al (1987) Failure of T3 to potentiate tricyclic antidepressant response. *Journal of Affective Disorders*, **13**, 267–272.

Hamilton, M. (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, **23**, 56–62.

Joffe, R. T. & Singer, W. (1990) A comparison of triiodothyronine and thyroxine in the potentiation of

- tricyclic antidepressants. *Psychiatry Research*, **32**, 241–251.
- & — (1992) Thyroid hormone use to enhance the effects of drugs. *Clinical Neuropharmacology*, **15** (suppl. 1), 389A–390A.
- , —, Levitt, A. J., et al (1993) A placebo-controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. *Archives of General Psychiatry*, **50**, 387–393.
- Klaiber, E. L., Broverman, D. M., Vogel, W., et al (1979) Estrogen therapy for severe persistent depressions in women. *Archives of General Psychiatry*, **36**, 550–554.
- LeLorier, J., Gregoire, G., Benhaddad, A., et al (1997) Discrepancies between meta-analyses and subsequent large randomized controlled trials. *New England Journal of Medicine*, **337**, 536–542.
- Lewis, G., Churchill, R. & Hotopf, M. (1997) Systematic reviews and meta-analysis. *Psychological Medicine*, **27**, 3–7.
- Maes, M., Vandoolaeghe, E. & Desnyder, R. (1996) Efficacy of treatment with trazodone in combination with pindolol or fluoxetine in major depression. *Journal of Affective Disorders*, **41**, 201–210.
- Malison, R. T., Anand, A., Pelton, G. H., et al (1999) Limited efficacy of ketoconazole in treatment-refractory major depression. *Journal of Clinical Psychopharmacology*, **19**, 466–470.
- Moher, D., Pham, B., Jones, A., et al (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, **352**, 609–613.
- Moreno, F. A., Gelenberg, A. J., Bachar, K., et al (1997) Pindolol augmentation of treatment-resistant depressed patients. *Journal of Clinical Psychiatry*, **58**, 437–439.
- Nierenberg, A. A. & White, K. (1990) What next? A review of pharmacologic strategies for treatment resistant depression. *Psychopharmacology Bulletin*, **26**, 429–460.
- Nolen, W. A., Haffmans, P. M., Bouvy, P. F., et al (1993) Monoamine oxidase inhibitors in resistant major depression. A double-blind comparison of brofaromine and tranylcypromine in patients resistant to tricyclic antidepressants. *Journal of Affective Disorders*, **28**, 189–197.
- Perez, V., Soler, J., Puigdemont, D., et al (1999) A double-blind, randomized, placebo-controlled trial of pindolol augmentation in depressive patients resistant to serotonin reuptake inhibitors. *Archives of General Psychiatry*, **56**, 375–379.
- Poirier, M. F. & Boyer, P. (1999) Venlafaxine and paroxetine in treatment-resistant depression. Double-blind, randomised comparison. *Psychiatry*, **175**, 12–16.
- Rybakowski, J. K., Suwalska, A. & Chlopocka-Wozniak, M. (1999) Potentiation of antidepressants with lithium or carbamazepine in treatment-resistant depression. *Neuropsychobiology*, **40**, 134–139.
- Sackett, D. L. & Cook, R. J. (1995) The number needed to treat: a clinically useful measure of treatment effect. *BMJ*, **310**, 452–454.
- Shelton, R. C., Tollefson, G. D., Tohen, M., et al (2001) A novel augmentation strategy for treating resistant major depression. *American Journal of Psychiatry*, **158**, 131–134.
- Shergill, S. S. & Katona, C. L. E. (1996) Pharmacological choices after one antidepressant fails: a survey of UK psychiatrists. *Journal of Affective Disorders*, **43**, 19–25.
- StataCorp (1999) *Stata Statistical Software*. College Station, Texas: Stata Corporation.
- Stimpson, N., Lewis, G. & Agrawal, N. (2000) Pharmacological versus non-pharmacological interventions for refractory depression (Protocol for a Cochrane Review). *Cochrane Library*, issue 4. Oxford: Update Software.
- Thase, M. E. & Rush, A. J. (1995) Treatment-resistant depression. In *Psychopharmacology: The Fourth Generation of Progress* (eds F. E. Bloom & D. J. Kupfer), pp. 1081–1097. New York: Raven Press.
- Tyrer, P., Marsden, C. A., Casey, P., et al (1987) Clinical efficacy of paroxetine in resistant depression. *Journal of Psychopharmacology*, **1**, 251–257.
- World Psychiatric Association (1974) Symposium on therapy resistant depression. *Pharmacopsychiatry*, **7**, 69–224.
- Zohar, J., Shapira, B., Oppenheim, G., et al (1985) Addition of estrogen to imipramine in female resistant depressives. *Psychopharmacology Bulletin*, **21**, 705–706.
- Zusky, P. M., Biederman, J., Rosenbaum, J. F., et al (1988) Adjunct low dose lithium carbonate in treatment-resistant depression: a placebo-controlled study. *Journal of Clinical Psychopharmacology*, **8**, 120–124.