Recent placebo-controlled acute trials in bipolar depression: focus on methodology

David J. Muzina¹ and Joseph R. Calabrese²
¹ Cleveland Clinic Foundation, Cleveland, Ohio, USA
² Department of Psychiatry, Case Western Reserve University School of Medicine, University Hospitals of Cleveland, Cleveland, Ohio, USA

Abstract

The completion of three recent large-scale, double-blind controlled acute trials in bipolar I depression has improved our understanding of the management of major depressive episodes associated with bipolar disorder. In contrast to the cross-over designs used in the early studies of lithium in bipolar depression, the designs utilized in these recent studies have employed random assignment to parallel arms including the use of placebo as a monotherapy in one study. The analyses of recent studies have all been conducted on intent-to-treat data, and included two types, change from baseline analyses and responder analyses. Lamotrigine monotherapy was shown to be superior to placebo with both types of analyses on the Montgomery–Asberg Depression Rating Scale (MADRS) and the Clinical Global Impressions (CGI) scales, but not the 17-item Hamilton Depression Rating Scale (HAMD) (n = 195). The percentage of patients responding to placebo as a monotherapy were 29, 26 and 37%, respectively; there were no differences in switch rates (5% vs. 5%). Paroxetine augmentation was no better than placebo augmentation overall with both types analyses on the CGI and HAMD (n = 117); the MADRS was not used. In patients with lithium levels ≤0.8 mequiv./l, the change from baseline analysis showed paroxetine to be superior to placebo, but responder analyses were negative; switch rates with paroxetine, imipramine, and placebo were 0, 8 and 2%. Moclobemide monotherapy was similar in efficacy to imipramine (n = 156), but had a lower rate of switching (4% vs. 11%).

Received 28 July 2002; Reviewed 29 October 2002; Revised 29 January 2003; Accepted 9 February 2003

Key words: Bipolar depression, lamotrigine, placebo, paroxetine, tricyclic antidepressants.

Early bipolar depression studies

Until very recently, no large-scale, placebo-controlled, double-blind controlled trials had been carried out in bipolar I depression. The controlled trials conducted in bipolar I depression between 1968 and 1978 compared the acute antidepressant effects of lithium monotherapy to placebo, but did not employ random assignment to parallel groups (Baron et al., 1975; Donnelly et al., 1978; Fieve et al., 1968; Goodwin and Murphy, 1969; Goodwin et al., 1972; Greenspan et al., 1970; Mendels, 1976; Noyes et al., 1974; Stokes et al., 1971). Although these early classic studies suggest at least moderate efficacy of lithium, methodological problems limit the interpretation of these data. These studies did not exclusively enrol patients with bipolar disorder, nor did they report switch rates into hypomania, mania, or mixed states. Typically, they did not select primary outcome analyses ‘a priori’. Instead, a variety of outcome measures were selected based upon the phenomenology of the illness and the consensus of scientific opinion. In contrast to early studies, most recent studies have attempted to satisfy the regulatory requirement of selecting a primary outcome measure and then attempting to power a study to show significant differences on this measure. Most of these early analyses were also conducted on observed data and none employed the more conservative intent-to-treat analysis with last observations carried forward (LOCF). In addition, the use of lithium/placebo cross-over designs may have confounded early estimates of lithium’s antidepressant efficacy (Faedda et al., 1993). Recently, the completion of several large-scale, double-blind controlled trials in bipolar depression has informed the treatment of the depressed phase of bipolar disorder (Calabrese et al., 1999a; Nemeroff et al., 2001; Silverstone et al., 2001).
Ten studies involving a total of 533 patients have explored the clinical efficacy of various marketed antidepressants in the management of bipolar depression between 1968 and 2001 (see Table 1: Baumhackl et al., 1989; Cohn et al., 1989; Fieve et al., 1968; Himmelhoch et al., 1991; Kessel and Holt, 1975; Nemeroff et al., 2001; Sachs et al., 1994; Silverstone et al., 2001; Thase et al., 1992; Watanabe et al., 1975). None of these studies examined the spontaneous switch rate associated with the natural course of the illness through the administration of placebo as a monotherapy (see Calabrese et al., 1999c for a detailed review of switch rates reported in these early studies). In contrast to the lithium studies in bipolar depression, which uniformly employed cross-over designs, all of these more recent studies used random assignment to parallel groups. Five of the studies limited enrolment to patients with bipolar disorder and four examined mixed populations of both unipolar and bipolar disorder. One of the five studies enrolled patients with bipolar I or II disorder, but did not stratify for this variable (Himmelhoch et al., 1991). One study used placebo during a 2- to 4-wk lead-in (Fieve et al., 1968) and one study employed random assignment to a quasi-parallel placebo group (Cohn et al., 1989). Four studies permitted the use of mood stabilizers (Cohn et al., 1989; Nemeroff et al., 2001; Sachs et al., 1994; Watanabe et al., 1975), but only two standardized their use (Nemeroff et al., 2001; Sachs et al., 1994). Responder analyses were usually carried out on observed data only, but in one study that lacked a placebo group, Himmelhoch et al. (1991) did employ the more conservative intent-to-treat analysis. The recent augmentation study of Nemeroff et al. (2001) also employed random assignment to parallel arms followed by intent-to-treat analyses.

Of the early studies, five allow for the following tentative conclusions: (1) Lithium was less effective than imipramine in ameliorating the symptoms of bipolar depression (Fieve et al., 1968). (2) Fluoxetine was superior in efficacy to both imipramine and quasi-placebo (86% vs. 57% vs. 38% of patients responded, respectively). Although switch rates favoured fluoxetine, the protocol did not specify the method by

---

**Table 1. Randomized trials of antidepressant treatments for bipolar depression**

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Subjects (duration)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fieve et al.</td>
<td>Random assignment to parallel</td>
<td>$n = 29$ (3 wk)</td>
<td>Both lithium and imipramine more effective than placebo. Greater decrease in depression for imipramine group (58%) than lithium group (32%) (placebo = 0%, lead-in)</td>
</tr>
<tr>
<td>(1968)</td>
<td>monotherapy groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kessel et al.</td>
<td>Random assignment to mixed</td>
<td>$n = 14$ (6 wk)</td>
<td>Maprotiline response rate of 66% compared to imipramine 40%</td>
</tr>
<tr>
<td>(1975)</td>
<td>parallel monotherapy groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watanabe et al.</td>
<td>Random assignment to parallel</td>
<td>$n = 5$ (5 wk)</td>
<td>Lithium and imipramine equally effective</td>
</tr>
<tr>
<td>(1975)</td>
<td>monotherapy groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohn et al.</td>
<td>Random assignment to mixed</td>
<td>$n = 89$ (6 wk)</td>
<td>Greater response rates observed with fluoxetine (86%) and imipramine (57%) compared to placebo (38%)</td>
</tr>
<tr>
<td>(1989)</td>
<td>parallel groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baumhackl et</td>
<td>Random assignment to mixed</td>
<td>$n = 32$ (4 wk)</td>
<td>Antidepressant response observed in 53% moclobemide-treated group and 60% in imipramine group</td>
</tr>
<tr>
<td>(1989)</td>
<td>monotherapy groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Himmelhoch et</td>
<td>Random assignment to parallel</td>
<td>$n = 56$ (6 wk)</td>
<td>Tranylcypromine more effective than imipramine (89% vs. 48%)</td>
</tr>
<tr>
<td>et al. (1991)</td>
<td>monotherapy groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thase et al.</td>
<td>Double-blind cross-over</td>
<td>$n = 16$ (4-8 wk)</td>
<td>Tranylcypromine effective in 75% patients who had failed imipramine</td>
</tr>
<tr>
<td>(1992)</td>
<td>follow-up study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sachs et al.</td>
<td>Double-blind, random assignment</td>
<td>$n = 19$ (8 wk)</td>
<td>Bupropion and desipramine possessed similar antidepressant efficacy (63% vs. 71%)</td>
</tr>
<tr>
<td>(1994)</td>
<td>to parallel augmentation groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-yr extension)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silverstone et al.</td>
<td>Random assignment to parallel</td>
<td>$n = 156$ (8 wk)</td>
<td>No differences in efficacy were detected between moclobemide and imipramine</td>
</tr>
<tr>
<td>(2001)</td>
<td>groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nemeroff et al.</td>
<td>Random assignment to mixed</td>
<td>$n = 117$ (10 wk)</td>
<td>Paroxetine and imipramine were superior to placebo for patients with low lithium levels (&lt;0.8)</td>
</tr>
<tr>
<td>(2001)</td>
<td>parallel groups</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
which patients were described as having switched and it is likely that switches into hypomania were ignored (Cohn et al., 1989). (3) Tranylcypromine monotherapy was superior in efficacy to imipramine in the treatment of bipolar depression (81% vs. 48% of patients responded, respectively) and switch rates (defined as the rate of induction of manic episodes meeting research diagnostic criteria) for both drugs were relatively high and similar (21% vs. 25% of patients, respectively) (Himmelhoch et al., 1991). This higher rate of switching may have been due to the use of these drugs without lithium. (4) Bupropion was similar in efficacy to desipramine in bipolar depression (63% vs. 71% of patients responded, respectively), but switch rates (defined as the rate of induction of manic episodes meeting DSM-IV criteria) favoured bupropion (desipramine, 50% vs. bupropion, 11%) (Sachs et al., 1994).

Recent large-scale, double-blind-controlled trials

The completion of three large-scale, double-blind controlled acute trials in bipolar I depression has substantially improved our understanding of the clinical management of the depressed phase of bipolar disorder (see Table 2: Calabrese et al., 1999a; Nemeroff et al., 2001; Silverstone et al., 2001).

In the first study of bipolar depression to employ random assignment to parallel monotherapy groups including placebo as a monotherapy, Calabrese et al. (1999a) performed a 7-wk outpatient multicentre trial, which compared the efficacy of two doses of lamotrigine monotherapy [50 mg (n = 66) and 200 mg (n = 63)] with placebo monotherapy (n = 66). Patients with a history of rapid cycling within the last 12 months were excluded. Psychiatric evaluations, including the Hamilton Rating Scale for Depression (HAMD), the Montgomery–Asberg Depression Rating Scale (MADRS), Mania Rating Scale (MRS), and the Clinical Global Impressions Scale for Severity (CGI-S) and Improvement (CGI-I), were completed at each weekly visit. A dose of 200 mg/d lamotrigine demonstrated antidepressant efficacy on the 17-item HAMD, HAMD-Item 1, MADRS, CGI-S, and CGI-I compared to placebo. Improvements were seen as early as week 3, which corresponds to 1 wk at 50 mg/d lamotrigine (see Figure 1). Statistically significant differences were seen on both observed data and intent-to-treat data for all measures, except the HAMD, which was only positive on observed data. Lamotrigine at 50 mg/d also demonstrated efficacy compared to placebo on several measures. The proportions of patients exhibiting a response on CGI-I were 51, 41 and 26% for 200 mg/d lamotrigine, 50 mg/d lamotrigine and placebo groups, respectively (see Figure 2). Adverse events and other safety results were similar across treatment groups, except for a higher rate of headache in the lamotrigine groups. Switching was no more prevalent in the lamotrigine-treated patients than in the placebo-treated patients (5.4% vs. 4.6% of patients, respectively).

A strategic issue in the design of a bipolar depression study is whether to develop the new compound as a monotherapy or for use in augmentation. This decision is based upon the availability of preliminary data. In order to demonstrate the safety of a monotherapy comparison, there must be some preliminary data regarding the mood-stabilizing properties of the drug. This is viewed as being necessary to avoid unacceptably high rates of switching, which have been observed in prior monotherapy

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Subjects (duration)</th>
<th>Results</th>
<th>Switch rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calabrese et al. (1999a)</td>
<td>Random assignment to parallel monotherapy groups including placebo</td>
<td>n = 195 (7 wk)</td>
<td>Both 50 and 200 mg doses of lamotrigine more effective than placebo</td>
<td>Lamotrigine = 5.4% Placebo = 4.6%</td>
</tr>
<tr>
<td>Nemeroff et al. (2001)</td>
<td>Random assignment to mixed parallel groups including placebo</td>
<td>n = 117 (10 wk)</td>
<td>Paroxetine and imipramine were superior to placebo for patients with low lithium levels (&lt;0.8)</td>
<td>Paroxetine + lithium = 0% Imipramine + lithium = 7.7% Placebo + lithium = 2.3%</td>
</tr>
<tr>
<td>Silverstone et al. (2001)</td>
<td>Random assignment to parallel groups; no placebo; mood stabilizers permitted</td>
<td>n = 156 (8 wk)</td>
<td>No differences in efficacy were detected between moclobemide and imipramine</td>
<td>Moclobemide = 9.3% Imipramine = 13%</td>
</tr>
</tbody>
</table>
Early naturalistic observations regarding the mood-stabilizing properties of lamotrigine, which included patients treated with monotherapy, allowed for the design of this monotherapy comparison (Calabrese et al., 1999b,c). The latter study generated data which support the conclusion that the spontaneous rate of cycling during the first 7 wk of treatment in patients with bipolar I disorder is approx. 5% if patients with a recent history of rapid cycling are excluded.

More recently, investigators have begun to examine the antidepressant efficacy of the selective serotonin reuptake inhibitors when added to lithium and/or valproate and/or carbamazepine in bipolar depression. Also employing random assignment to parallel groups, Nemeroff et al. (2001) recently reported on an 8-wk multicentre trial, which compared the efficacy of moclobemide (450–750 mg/d) (n = 54) with imipramine (150–250 mg/d) (n = 54). Lithium was being taken by 46% of the patients in the moclobemide group of whom 2 were also on carbamazepine and 49% in the imipramine group of whom 5 were also taking carbamazepine and 1 valproate. Ten patients in the moclobemide group were taking carbamazepine as their sole mood stabilizer and another patient was on valproate; the corresponding figures for the imipramine group were 9 on carbamazepine and 2 on valproate. Although similar antidepressant efficacy was observed between the two groups (69% vs. 74% of patients responded, respectively), fewer side-effects were associated with moclobemide.
(anticholinergic and weight gain) and fewer switches were noted with moclobemide. The superiority of moclobemide over imipramine in switch rates was observed, not only when switching was defined as an increase in the total Young Mania Rating Scale (YMRS) score to \(\geq 10\) (moclobemide, 9.3% patients vs. imipramine, 13% patients), but also when emergence of manic symptoms was sufficient to require drug discontinuation (moclobemide, 3.7% patients vs. imipramine 11.1% patients). These rates of switching may have been low because of the concurrent use of mood stabilizers.

Discussion

Recently conducted bipolar depression studies now routinely employ random assignment to placebo-controlled monotherapy comparisons analysed with LOCF. In contrast to early bipolar depression research, these studies tend to be sponsored by the pharmaceutical industry. Lamotrigine has been shown to be superior to placebo, paroxetine augmentation has been shown to be better than placebo but only in patients with lower levels of lithium, and moclobemide was similar in efficacy to imipramine. Previously reported high rates of switching (up to 67%) have not been replicated in recent controlled trials, but these recent studies do confirm the prior impression that the tricyclic antidepressants lead to the highest risk of switching. The lower switch rates observed with lamotrigine, paroxetine and moclobemide indicate that these drugs have a special application in the management of patients prone to antidepressant-induced switching (i.e. patients who have a history of breakthrough hypomanic/mania on their current therapy and those with rapid cycling). Of the recently studied medications, lamotrigine was administered without concomitant lithium or valproate. The lamotrigine study also provided the first information on the spontaneous rate of the development of hypomania or mania during the acute treatment of the depressed phase of bipolar I disorder as reflected by switch rates on placebo monotherapy (Calabrese et al., 1999a). Data from this study indicated that patients with active forms of bipolar I depression spontaneously switch at a rate of 4–5% over the first 7 wk of pharmacotherapy (Calabrese et al., 1999a).

These recently designed bipolar depression studies represent substantial progress when compared to the studies designed between 1960 and 1980. However, it is likely that these recent studies will also be viewed by the next generation of bipolar depression research as being less than ideal. There remains substantial unmet need in bipolar depression and some of the recent findings will be difficult to generalize.

The consensus of opinion regarding the definition of ‘acute response’ in this recent literature is that ‘magnitude of improvement reflected by a 50% decrease in baseline symptom severity (total scores) at the time of study end-point with last observations carried forward’. Although there is good consensus that this definition optimally separates treatment arms from placebo, we do not understand why higher degrees of improvement, as manifested by higher percentage decreases in baseline symptom severity, do not enhance separation from placebo. In fact, these post-hoc analyses suggest definitions of response that require higher degrees of improvement perform less well. Definitions using 60–80% decreases in baseline symptom severity do not separate from placebo. We believe this reflects a substantial, and somewhat obscured, unmet need for patients experiencing a major depression. If we are to be scientifically rigorous, we must conclude that the currently available medications

\[
\begin{array}{c|c|c|c}
\text{Placebo} & \text{LTG 50} & \text{LTG 200} \\
\hline
\text{HAMD-17} & 37 & 45 & 51 \\
\text{MADRS} & 29 & 48 & 54 \\
\text{CGI-I} & 26 & 41 & 51 \\
\end{array}
\]

\(\text{Figure 2. Percentage of patients showing a response to treatment at end-point.} \ a \text{Response defined as } \geq 50\text{% reduction on the 17-item HAMD or MADRS scales or a rating of very much improved or much improved on the CGI-I Scale.} \ b \ p < 0.05 \text{ vs. placebo; } c \ p < 0.1 \text{ vs. placebo.}
\]
are not sufficiently effective, even when used for 2–3 months. The current definition of response allows a patient who enters a study with markedly severe illness to be called an acute ‘responder’ at the end of 7–10 wk of treatment if the magnitude of their symptomatology decreases to only moderate severity. This definition of acute ‘response’ performs well from a statistical perspective, but obscures unmet need and does not address substantial human suffering.

These recently conducted bipolar depression studies do not report remission rates. This is in some contrast to recently conducted acute olanzapine mania studies (Tohen et al., 2000) for which there is emerging consensus that ‘remission’ should be defined as a total score on a symptom severity scale that has been shown to be the lower limits of normal in a general population of normal volunteers, i.e. a score of 10 on the YMRS. There is no consensus among mania investigators whether this should be required over one time-point reflecting just the past 7 d or two separate time-points reflecting the last 2 wk; the olanzapine studies only required one time-point for remission. It is unclear if the more rigorous definition requiring the second time-point improves separation from placebo. Regardless, we believe using one time point obscures unmet need.

We believe there are some methodological features of recently designed bipolar depression studies that will eventually be viewed as limiting generalizability. The exclusion of patients with rapid-cycling bipolar disorder who otherwise meet diagnostic criteria for the disorder tends to artificially suppress the rate of switching/cycling observed on placebo as well as the rate of switching observed with the experimental agent (Calabrese et al., 1999a; Nemeroff et al., 2001). This methodological convention is statistically and scientifically sound. However, it tends to obscure the extent to which these drugs can destabilize the course of bipolar disorder and limits the generalizability of the findings. These new agents are going to be frequently used in the management of patients with rapid cycling because the frequent recurrence of treatment-refractory depression is emerging as the hallmark of rapid-cycling bipolar disorder (Calabrese et al., 2001).

There is also little or no consensus in the literature regarding the definition of switching. As a result, most recent large-scale studies employ different definitions of switching and this practice will make it difficult to compare data from different trials. We have previously proposed that switching be defined as the rate at which patients cycle from the depressed phase of the illness into the hypomanic, manic, or mixed phase meeting diagnostic criteria during the first 6–8 wk of treatment (Calabrese et al., 1999c). Since there is little consensus over what is the optimal definition of switching, there may be a rationale to collect data using all of the various measures reported in prior studies. These would include the most specific and least sensitive definitions such as switching to a manic state severe enough to result in hospitalization, or symptoms severe enough to justify premature study drug discontinuation. Measures of moderate sensitivity and moderate specificity would include those defining mania or hypomania according to a specific diagnostic nomenclature, such as the DSM-IV. Measures with the most sensitivity but the least specificity would include those recording switching as an adverse event at the discretion of the investigator, or those reporting total scores on mania rating scales exceeding certain thresholds. For example, a YMRS total score of 12.5 has been described as reflecting the median euthymic rating, 15 as the lower limits of mania, 20 as mania of moderate severity and 25 as the lower limit of mania of sufficient severity to require hospitalization (Young et al., 1978).

Switches into full manic or hypomanic episodes are likely to occur at low rates during acute treatment trials of only 6–10 wk duration. Monitoring for switching to mood states above the baseline is needed beyond the acute treatment setting. In addition, patients may experience an increase in mood lability or ‘roughening’ which should also be considered mood destabilization. Patients enrolled in outpatient studies are generally less severely ill and these factors must be considered methodologically in order to make future studies more generalizable.

In conclusion, there is emerging consensus that the greatest unmet need in bipolar disorder is the management of depressive symptoms and the major depressive episodes associated with the disorder. This conclusion is supported by the methods employed to develop these compounds and the virtual absence of drugs that have received regulatory approval for use in the treatment of the depressed phase of bipolar disorder. There exists an urgent need to develop mood stabilizers that stabilize mood ‘from below baseline’ (Ketter and Calabrese, 2002) by exerting marked antidepressant effects without destabilizing the overall course of bipolar disorder.

References


Calabrese JR, Rapport DJ, Kimmel SE, Shelton MD (1999c). Controlled trials in bipolar I depression: focus on switch rates and efficacy. European Neuropsychopharmacology 9 (Suppl. 4), S109–S112.


