TECHNICAL COMMENTARY

Treatments for bipolar II disorder

Introduction

Bipolar II disorder is a common, recurrent, and disabling psychiatric illness. The lifetime incidence ranges from 1% to 11% depending on the method of diagnosis. DSM-IV defines bipolar II disorder as a lifetime history of at least one episode of major depression plus at least one episode of hypomania. By definition, individuals with bipolar II disorder never experience full-blown mania, unlike those with a diagnosis of bipolar I disorder. Bipolar II disorder is characterised by multiple and often protracted depressive episodes, and a lower probability of returning to premorbid levels of functioning between episodes[1].

Method

We have included only systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) published in full text, in English, from the year 2010 that report results for people with a diagnosis of bipolar or related disorders. Due to the high volume of systematic reviews we have now limited inclusion to systematic meta-analyses. Where no systematic meta-analysis exists for a topic, systematic reviews without meta-analysis are included for that topic. Reviews were identified by searching the databases MEDLINE, EMBASE, and PsycINFO. Hand searching reference lists of identified reviews was also conducted. When multiple copies of review topics were found, the most recent and/or comprehensive review was included.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist that describes a preferred way to present a meta-analysis[2]. Reviews with less than 50% of items checked have been excluded from the library. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomised controlled trials (RCTs) may be downgraded to moderate or low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large, there is a dose dependent response or if results are reasonably consistent, precise and direct with low associated risks (see end of table for an explanation of these terms)[3]. The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of NeuRA (Neuroscience Research Australia).

Results

We found one systematic review that met inclusion criteria[1].

- Moderate to low quality evidence suggests quetiapine may be an effective treatment for bipolar II depression.
- Low quality evidence is uncertain as to the benefits of other medications.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Swartz HA, Thase ME

Pharmacotherapy for the treatment of acute bipolar II depression: current evidence

Journal of Clinical Psychiatry 2011 72: 356-66

View review abstract online

Comparison	**All pharmaceutical treatments for bipolar II depression.**
Summary of evidence	**Moderate to low quality evidence (medium-sized sample, unable to assess consistency or precision, direct) suggests quetiapine may be an effective treatment for bipolar II depression.** **Low quality evidence (small samples, and/or samples with mixed bipolar I and bipolar II disorders) is uncertain as to the benefits of other medications.**

Depression

8 weeks of quetiapine vs. placebo

2 RCTs, N = 361, ES = 0.45 (300 mg/d) or 0.54 (600 mg/d), remission rates were 39.3 %, 37.7 % and 20.4% for quetiapine 300 mg/d, 600 mg/d, and placebo.

7-10 weeks of lamotrigine vs. placebo

5 RCTs, N = 305, lamotrigine did not differ significantly from placebo.

8 weeks of lithium plus lamotrigine vs. lithium plus placebo

1 RCT, N = 124, significantly more patients responded to adjunctive lamotrigine (51.6%) than placebo (31.7%), however the sample also included bipolar 1.

6 weeks of lamotrigine vs. gabapentin vs. placebo

1 crossover RCT, N = 31, response rate for lamotrigine was 52%, which was superior to gabapentin (26%) and placebo (23%), however the sample also included bipolar 1 and unipolar depression.

16 weeks of lamotrigine vs. inositol vs. risperidone

1 RCT, N = 47, lamotrigine may be superior to risperidone with inositol showing an intermediate effect, however the sample also included bipolar 1.

16 weeks of lamotrigine vs. lithium

1 open-label RCT, N = 98, both groups showed significant improvement on depression scores over time with no differences between groups. 76% of the lamotrigine group and 59% of the lithium group met criteria for remission without switch into hypomania.

12 weeks of lamotrigine vs. citalopram

1 RCT, N = 20, both groups showed clinically significant improvement on depression scores with no

statistically significant differences between groups, however the sample also included bipolar 1. <u>12 weeks of lithium vs. venlafaxine</u> 1 open-label RCT, N = 83, venlafaxine was superior to lithium, even among the subset of patients with a history of rapid cycling, on measures of depressive symptoms as well as proportions responding and remitting. Rates of treatment-emergent affective symptoms were low and comparable between groups. <u>6 weeks of divalproex vs. placebo</u> 1 RCT, N = 18, divalproex was superior to placebo on measures of depressive symptoms, however the sample also included bipolar 1. <u>26 weeks of bupropion or paroxetine vs. placebo</u> 1 RCT, N = 355, no difference in response rates between groups receiving antidepressants and placebo. Rates of treatment-emergent affective switches did not differ between groups. <u>6 weeks of paroxetine vs. lithium or divalproex</u> 1 RCT, N = 28, both groups improved over time, but there were no significant differences between groups, and the sample also included bipolar 1. <u>10 weeks of sertraline vs. bupropion vs. venlafaxine</u> 1 RCT, N = 159, overall response rates ranged from 43–55% and did not differ significantly among groups. The sample also included bipolar 1. <u>9 months of escitalopram vs. placebo</u> 1 crossover RCT, N = 10, escitalopram was associated with significant reductions in depression severity, percentage of days depressed or high, and impairment. <u>6 weeks of pramipexole vs. placebo</u> 1 RCT, N = 21, pramipexole was associated with greater reductions in depression scores and no greater rates of switching. <u>6 weeks of modafinil vs. placebo</u> 1 RCT, N = 86, there were no differences in response rates between groups within the bipolar II subgroup. <u>4 months of ethyl-eicosapentanoate vs. placebo</u> 1 RCT, N = 120, no differences in outcome measures across treatment groups in the entire sample. The sample also included bipolar 1. <u>12 weeks of ethyl-eicosapentanoate 1 g/d or 2 g/d vs. placebo</u> 1 RCT, N = 76, there were statistically significant greater symptomatic improvement with ethyl-eicosapentanoate (both doses), however the sample also included bipolar 1.	
Consistency‡	No measure of consistency is reported.
Precision§	No measure of precision is reported.
Directness‖	Direct

Explanation of acronyms

DSM = American Psychiatric Association's Diagnostic and Statistical Manual, N = number of participants, vs. = versus

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Explanation of technical terms

* Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results; publication bias - trials that are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small[4].

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives that are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives that are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomised trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardsed mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) that allows results from different scales to be combined and compared. Each study's mean difference is then given a weighting depending on the size of the sample and the variability in the data. 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 and over represents a large effect[4].

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, a RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. A RR or OR of 1.00 means there is no difference between groups. A medium effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.2[5]. lnOR stands for logarithmic OR where a lnOR of 0 shows no difference between groups. Hazard ratios measure the effect of an explanatory variable on the hazard or risk of an event.

Correlation coefficients (eg, r) indicate the strength of association or relationship between variables. They can provide an indirect indication of prediction, but do not

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

confirm causality due to possible and often unforseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association. Unstandardised (*b*) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the independent variable, statistically controlling for the other independent variables. Standardised regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of effect across studies (i.e. heterogeneity or variability in results) that is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. I^2 is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent considerable heterogeneity and over this is considerable heterogeneity. I^2 can be calculated from Q (chi-square) for the test of heterogeneity with the following formula[4];

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

§ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data (standardised mean differences, not weighted mean differences) is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, these criteria should be relaxed[6].

‖ Indirectness of comparison occurs when a comparison of intervention A versus B is not available but A was compared with C and B was compared with C that allows indirect comparisons of the magnitude of effect of A versus B. Indirectness of population, comparator and/or outcome can also occur when the available evidence regarding a particular population, intervention, comparator, or outcome is not available and is therefore inferred from available evidence. These inferred treatment effect sizes are of lower quality than those gained from head-to-head comparisons of A and B.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

References

1. Swartz HA, Thase ME (2011): Pharmacotherapy for the treatment of acute bipolar II depression: current evidence. *Journal of Clinical Psychiatry* 72: 356-66.
2. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMAGroup (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 151: 264-9.
3. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
4. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
5. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
6. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 32 for Windows*.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au