Abstract

Background: Psychological therapies may play an important role in the treatment of bipolar disorders. Several meta-analyses that examine the effectiveness of psychotherapies for patients with bipolar disorder include conclusions about the impact upon bipolar depression. However, these tend not to consider differences in depression outcome depending upon whether the therapy primarily targets acute depression, nor severity of baseline depression. This may affect the conclusions drawn about the effectiveness of these therapies for acute bipolar depression treatment.

Objectives: This meta-analysis explored the effectiveness of psychological therapies in reducing bipolar depression, in particular examining whether: (1) the effect of therapy is greater when baseline depressive symptoms are more severe, and (2) the effect of therapy is greater when the primary focus of the therapy is the treatment of acute bipolar depression?

Data sources: A systematic search was conducted using the following electronic databases; Cochrane Controlled Register of Trials (1996), MEDLINE (1966 onwards), EMBASE (1980 onwards), PsycINFO (1974 onwards), Scopus, Web of Science and Clinical Trials Registries (listed at:https://www.hhs.gov/ohrp/international/clinical-trial-registries/index.html).

Eligibility criteria: Eligible studies were randomized controlled trials evaluating a psychological intervention for adults diagnosed with Bipolar I or II disorder. The comparators were usual care, wait-list, placebo, active treatment control. Post-treatment depression status was required to be measured continuously using a validated self- or observer- report measure, or categorically by a validated diagnostic instrument or clinical diagnosis by a suitably qualified person.

Data extraction and synthesis: Titles and abstracts were screened, followed by full texts. Two reviewers conducted each stage until agreement was reached, and both independently extracted study information. Means, standard deviations (SDs) and number of participants were retrieved from articles and used to perform a meta-analysis. The primary outcome was depressive symptom score.

Results: The database search identified 6388 studies. After removing the duplicates, 3298 studies remained, of which, 28 studies were included in the qualitative review and 22 in the meta-analysis. Effect sizes range from -1.99 [-2.50, -1.49] to 0.89 [-0.12, 1.90]. There was low quality evidence of a significant effect on symptoms of depression for cognitive behavioural therapy and dialectical behaviour therapy. Trials of psychoeducation, mindfulness-based therapy, family therapy and interpersonal and social rhythm therapy showed no evidence of any effect on depression. We found no significant relationship between baseline depression score and depression outcome post-treatment when we controlled for therapy type and comparator. The result also showed that the effect sizes for studies targeting acute depression to be tightly clustered around a small overall effect size.

Conclusions: Some psychological therapies may reduce acute bipolar depression although this conclusion should be viewed with caution given the low quality of evidence. More research using similar therapy types and comparators is needed to better understand the relationship between depression status at baseline and outcome.

Key words: Bipolar depression, psychological interventions, meta-analysis
1. Introduction

Bipolar disorders are a category of chronic and recurrent psychiatric disorder with severe mood shifts from mania to depression. The disorders are associated with a shorter life expectancy (Laursen, 2011; Crump et al., 2013) and both psychiatric (Vieta et al., 2001; Krishnan et al., 2005) and medical (McIntyre et al., 2006; Kemp et al., 2010; Crump et al., 2013) comorbidity. Bipolar disorders affect about 1-2% of the population (Merikangas et al., 2011), with the majority of individuals experiencing multiple episodes of mania/hypomania and depression over a lifetime. Though mania and hypomania are particularly characteristic of bipolar disorder, the long-term course of a bipolar disorder is dominated by depressive symptoms (Judd et al., 2002; Perlis et al., 2006; Judd and Akiskal, 2003). These symptoms are associated with an increased risk of suicide attempts and completed suicide (Rihmer et al., 1990; Eroglu, Karakus & Tamam, 2013; Tidemalm et al., 2014), as well as comorbid anxiety in the form of panic attacks (Pini et al., 1997).

Psychological therapies may play an important role in the treatment of bipolar disorders. These include cognitive-behavioral therapy (CBT), psychoeducation, family focused therapy, interpersonal and social rhythms therapy (IPSRT), mindfulness-based cognitive therapy (MBCT), and dialectical behavior therapy (DBT). However, with respect to bipolar depression, findings from previous meta-analyses have been mixed. For example, Beynon et al. (2008) and Oud et al. (2016) reported that CBT could improve depressive symptoms, while Bi-Yu et al. (2016) did not find this. Similarly, in the case of psychoeducation, Oud et al. (2016) reported fewer depressive relapses following group therapy while Bond and Anderson (2015) found no effect. We propose that one possible explanation for this inconsistency is that the impact of therapy on depression may vary according to the primary target of the therapy protocol, such that therapy protocols that actively address depression as the target difficulty may be more likely to show an effect on depression than studies targeting a different difficulty. Furthermore, impact upon depression outcomes may be a function of baseline depression levels, because studies in which participants have higher levels of depression at baseline have greater potential to show an effect of the intervention upon depression.

Our aim was to answer whether psychological interventions for patients with Bipolar I or II Disorder reduce symptoms of depression, compared to usual care, wait-list, placebo or active control in randomized controlled trials, and whether intervention target, pre-treatment depression status, therapy modality, group versus individual therapy or psychological comorbidity at baseline affect treatment outcomes. We explored whether the effect was greater for studies including participants with higher levels of depression at baseline and for studies in which the therapy protocol primarily targeted acute depression as opposed to other difficulties. No other hypothesis were stated for the rest of the goals.

2. Methods

2.1. Protocol and registration

Our study protocol was registered on the PROSPERO database (CRD42019133442). Our report is written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline.

2.2. Eligibility criteria

Inclusion criteria were:
Randomised controlled trials (RCT) evaluating the efficacy or effectiveness of a psychological intervention for adults diagnosed with Bipolar I or II disorder using a validated diagnostic instrument (SCID, SADS, CIDI, PSE-10 SCAN, and MINI) or according to clinical diagnosis by a suitably qualified person. Psychological interventions were defined broadly: “interpersonal or informational activities, techniques, or strategies that target biological, behavioral, cognitive, emotional, interpersonal, social, or environmental factors with the aim of improving health functioning and well-being” (Darke, 2016). "Randomised" was defined as the study reporting use of random allocation of participants to conditions: we did not specify randomisation method required, however this considered within in risk of bias assessment. We did not place restrictions upon whether or not the psychological intervention should be adjunctive (i.e. delivered in addition to ongoing medication as part of usual care).

Studies using as a control condition one of the following: treatment / care as usual, wait-list, active control, and placebo. In keeping with standard definitions used by the Cochrane database (Faltinsen et al., 2019), usual care was defined as an intervention that reflects locally accepted treatment practices for individuals with bipolar disorder. Wait list was defined as receiving the experimental intervention after the trial ended. Active treatment defined as provision of a comparator psychological treatment that was not intended to replicate standard local care, nor to simply control for non-specific or shared factors. Psychological placebo was defined as an intervention that targets the nonspecific or shared components of psychological treatments.

Studies using as a treatment outcome depression symptom level assessed at pre and post-treatment, and measured either through researcher/clinician rating, using either a continuous or categorical scale.

Studies investigating participants aged 16 and above.

Studies published in English.

Studies published between 1952 and 2020.

2.3. Search strategy
The search strategy included terms for bipolar disorders (e.g., mania; manic depression), depression (e.g., depressive), therapy (e.g., psychotherapy; behav* activation), and randomized control trials (e.g., random allocation; randomisation). These were used to determine subject headings unique to each database. The following were searched up to an including end date of October 2020: Cochrane Controlled Register of Trials (1996), MEDLINE (1966 onwards), EMBASE (1980 onwards), PsycINFO (1974 onwards), Scopus, Web of Science and Clinical Trials Registries (listed at:https://www.hhs.gov/ohrp/international/clinical-trial-registries/index.html) (search terms for MEDLINE in the Supplement).

In addition, the reference lists of relevant systematic reviews and meta-analyses were examined for potentially eligible studies. We contacted protocol authors for further information when we were unable to find any corresponding outcome studies.
2.4. Study selection

All studies retrieved were compiled in EndNote software. After duplicates were removed both automatically and manually, two reviewers (SY, KB) independently screened a random set of 20% of titles/abstracts retrieved by the search strategy, at which point agreement between reviewers was calculated (Kappa = .82). Discrepancies between reviewers were discussed between the two reviewers and resolved, with input from a third reviewer (AH) as required. When agreement was reached, the remaining studies were screened by one reviewer (SY). Once the full list of titles/abstracts was screened, the full text of potentially eligible studies was retrieved. For those studies for which we could not find the full-text article, study authors were contacted to request the article. Fifteen out of 17 requested articles were provided from study authors. Two reviewers (SY, KB) screened full articles following the same procedure as previously. A third reviewer (AH) involved was involved when discrepancies occurred.

2.5. Data extraction

A data extraction sheet was developed and pilot tested by two reviewers (SY and AH) using one randomly selected study. Data from included studies were then extracted with the extraction form. Two reviewers (SY, KB) independently extracted the information for all studies included in the review. Discrepancies between reviewers were discussed between them and resolved. A third reviewer (AH) was involved when an agreement could not be reached. For those studies for which we could not extract necessary information, study authors were contacted to request it. Requested information was provided by 4 study authors out of 8.

The following data were extracted: authors, year of publication, sample size, country in which the trial was conducted, study criteria, primary diagnosis, psychiatric comorbid conditions, target participants’ age, study setting (i.e., clinical setting, general population), research design, number of arms, type of comparison group (i.e., usual care, placebo, wait list, and another active treatment), intervention target (i.e., acute depression, remission), depression status, person who delivered the intervention and control, number of individuals at baseline, length of follow-up, total number of drop outs from baseline to last follow-up, duration of treatment, type of psychological treatment (i.e., CBT, IPT, behavioural therapy, psychoeducation, psychodynamic therapy, family therapy, counselling, mindfulness therapy, other), treatment modality (self-help, individual face-to-face therapy, group therapy, online/e-supported), treatment outcome variables, treatment outcome measures, type of outcome (continuous/categorical) and for each treatment outcome and each assessment time point: number of cases, mean and standard deviation for primary and secondary outcomes, and number of cases and number of events for categorical variables.

We extracted data at two follow-up points: post-treatment and longer-term follow-up. Post-treatment depression outcome was defined as the point immediately following the end of the acute treatment phase, and no later than 3 months after the end of treatment. Longer term outcomes were defined as being between the post-treatment outcome point and 18 months after the end of the treatment phase.

2.6. Quality of Evidence

The quality of selected studies were assessed independently at study level by two reviewers (SY, MP) using the Cochrane Collaboration’s risk of bias tool (Higgins et al., 2011). A third reviewer (AH) involved was involved when discrepancies were found. The following
items were assessed: selection bias (random sequence generation and allocation concealment); performance bias (blinding of participants and personnel); detection bias (blinding of outcome assessment); attrition bias (incomplete outcome data); reporting bias (selective reporting); and other bias (intention to treat analysis and group similarity at baseline, checks of the training of the therapist, manualisation of the therapy and whether fidelity to the therapy method had been assessed through rating tapes of all or only a subset of sessions). A judgement about the risk of bias arising from each item was made for each study. The judgement could be: low, unclear, or high risk of bias.

The Grading of Recommendations Assessment, Development and Evaluation (GRADE, 2013) framework was used to assess the overall quality of evidence for each outcome where it was possible to estimate a pooled effect. The GRADE approach categorizes the levels of the quality as very low, low, moderate, and high by utilizing several domains. According to this framework, quality rating is reduced by the presence of study limitations, inconsistency, indirectness, imprecision, and publication/reporting bias. One reviewer (KW) with subject matter expertise graded the overall quality of evidence using the GRADE framework, with decisions checked by a second reviewer (SY).

2.7. Data synthesis

Data such as means, standard deviations (SDs) and number of participants were retrieved from articles and used to perform a meta-analysis. The primary outcome was depressive symptom score. A meta-analysis was conducted using RevMan software, with pooled results expressed as the standardized mean differences (SMD) and associated 95% confidence intervals (95% CI). Standardized mean differences were calculated as our effect-size measure, which is the difference in post-treatment means divided by the pooled standard deviation, with adjustment for small sample bias (i.e. weights). A random-effects model was used and assumed common heterogeneity across all comparisons.

To examine the presence of heterogeneity, we used the $I^2$ statistic, which estimates the percentage of outcome variability that can be attributed to heterogeneity across studies. An $I^2$ value of 0% denotes no observed heterogeneity, whereas, 25% is “low”, 50% is “moderate” and 75% is “high” heterogeneity (Higgins et al., 2003). Publication bias was assessed through use of Egger’s linear regression model to statistically test for funnel plot asymmetry.

Pre-specified sensitivity analyses and/or meta-regression were conducted using Open Meta-Analyst software. As primary sensitivity analyses, we examined the relationship between the following: i) intervention target (acute depression / relapse prevention / co-morbid condition [i.e. anxiety, substance use, cognitive functioning]) and depression outcome; ii) depression status at pre-treatment and outcome. To examine the impact of intervention target, we grouped studies as follows: i) those that primarily and explicitly set out to treat acute depression; ii) those that set out primarily to prevent relapse; iii) those that set out to improve quality of life (only studies targeting relapse prevention and quality of life were present in sufficient numbers to allow this sensitivity analysis). To examine the impact of depression status pre-treatment upon outcome, we operationalised depression status in two ways: i) whether or not current acute depression was a stated inclusion criterion of the study; ii) mean depression score for study participants at pre-treatment. Additional secondary sensitivity analysis explored the relationship between outcome and: i) modality of therapy ii) group versus individual therapy, iii) psychological comorbidity at baseline.
To reduce heterogeneity and to reflect that our extracted studies included a range of psychological therapy approaches, subgroup analyses were conducted by splitting interventions into psychoeducation, CBT, MBCT, other therapies (DBT, family therapy, and IPSRT) for post-treatment and follow up. In addition, further subgroup analyses were conducted by examining studies within each intervention type according to the control condition used.

3. Results
We identified a total of 6388 studies through database searches and reference lists of relevant systematic reviews and meta-analyses. Of these, we removed 3298 duplicates automatically and manually (Figure 1). A further 3078 studies were excluded after title and abstract screening and as they were not randomized control trials. We then assessed 220 studies in full-text review for eligibility and of these, we excluded 188 studies as they did not meet the inclusion criteria. The remaining 32 studies were therefore included in our qualitative review. Of these, 22 were included in the meta-analysis (figure 1). The remaining 10 trials could not be included as the required outcomes were neither available in the article, nor from study authors.

3.1. Study characteristics
Six types of psychological therapies were identified: CBT, psychoeducation, DBT, MBCT, family therapy, and IPSRT. Studies were conducted in 11 countries: Australia, USA, Iran, UK, China, Spain, Egypt, France, Taiwan and New Zealand. Recruitment was from out-patient (21) and in-patient settings (1). For 6 studies, the information provided relating to study settings was unclear. Eleven studies examined CBT (3 group therapy, 8 individual face to face therapy); 13 psychoeducation (9 group therapy, 2 individual face to face therapy, 2 Online / e supported); 1 DBT (group therapy); 2 MBCT (group therapy); 1 family therapy (group therapy); 1 IPSRT (other). Control groups consisted of treatment as usual (16), waiting list (2), placebo (8), active control (2). In each study, the total number of participants ranged from 20 to 304. Overall follow-up ranged from 0 to 60 months. Duration of acute treatment ranged from 4 to 36 months (see Table 1 for further details).
Figure 1. PRISMA - Flow Diagram from Record Identification to Study Inclusion

- Records identified through database searching (n = 6388)
- Additional records identified through other sources (n = 154)
- Records after duplicates removed (n = 3298)
- Records screened (n = 3298)
- Full-text articles assessed for eligibility (n = 220)
- Studies meeting eligibility criteria (n = 32)
- Information provided is redundant. (n = 4)
- Studies included in qualitative synthesis (n = 28)
- Studies included in the meta-analysis (n = 22)
- Records excluded (n = 3078)
- Full-text articles excluded (n = 188)
  - The study is not published in English (n = 1)
  - Not a field study (n = 21)
  - The results are not derived exclusively from adults (n = 14)
  - The sample does not include patients with a diagnosis of bipolar I or II disorder (n = 7)
  - The results are not derived exclusively from patients with bipolar I or II disorder (n = 23)
  - The study does not test a psychological intervention (n = 27)
    - It is not a randomized controlled trial (n = 13)
    - The psychological intervention is not compared to any of the following conditions: 1. usual care 2. wait list, 3. placebo, or 4. active control (n = 12)
    - The primary goal is testing the feasibility of the intervention (n = 12)
    - The study do not report symptoms of depression as a treatment outcome (n = 32)
    - The studies does not provide outcomes of symptoms of depression at pre- and post-treatment (n = 20)
Table 1 Summary of the characteristics of the included studies

<table>
<thead>
<tr>
<th>Studies, country</th>
<th>Primary diagnosis</th>
<th>Study setting</th>
<th>Primary intervention target</th>
<th>Depressio n status at baseline</th>
<th>Number of individuals randomized</th>
<th>Length of follow-up (in months)</th>
<th>Duration of acute treatment (in weeks)</th>
<th>Duration of treatment sessions (in minutes)</th>
<th>Number of treatment sessions</th>
<th>Treatmen t outcome measure</th>
<th>Treatment group</th>
<th>Treatment modality</th>
<th>Comparato r group</th>
</tr>
</thead>
<tbody>
<tr>
<td>D’Souza 2010, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>58</td>
<td>12</td>
<td>12</td>
<td>90</td>
<td>12</td>
<td>MADRS</td>
<td>Psychoeducatio n</td>
<td>Group therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Costa 2011, Brazil</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Unclear/Unspecified</td>
<td>Not depressed</td>
<td>41</td>
<td>Up to 6</td>
<td>14</td>
<td>120</td>
<td>14</td>
<td>BDI</td>
<td>CBT</td>
<td>Group therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Schmitz 2002, Texas</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not specified</td>
<td>46</td>
<td>N/A</td>
<td>12</td>
<td>60</td>
<td>16</td>
<td>BDI</td>
<td>CBT</td>
<td>Individual face to face therapy</td>
<td>Waiting list</td>
</tr>
<tr>
<td>Dijk 2013, Canada</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Unclear/Unspecified</td>
<td>Not specified</td>
<td>26</td>
<td>N/A</td>
<td>12</td>
<td>90</td>
<td>12</td>
<td>BDI-II</td>
<td>DBT</td>
<td>Group therapy</td>
<td>Waiting list</td>
</tr>
<tr>
<td>Deckersbach 2018, USA</td>
<td>Bipolar I disorder</td>
<td>Unclear/Unspecified</td>
<td>Acute depression</td>
<td>Depressed</td>
<td>32</td>
<td>4</td>
<td>20</td>
<td>50</td>
<td>18</td>
<td>HDRS</td>
<td>CBT</td>
<td>Individual face to face therapy</td>
<td>Placebo</td>
</tr>
<tr>
<td>Weiss 2007, USA</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Other</td>
<td>Not specified</td>
<td>62</td>
<td>3</td>
<td>20</td>
<td>60</td>
<td>20</td>
<td>HDRS</td>
<td>CBT</td>
<td>Group therapy</td>
<td>Active control</td>
</tr>
<tr>
<td>Perich 2013, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>95</td>
<td>12</td>
<td>8</td>
<td>120-150</td>
<td>8</td>
<td>MADRS</td>
<td>MBCT</td>
<td>Group therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Miklowitz 2000, USA</td>
<td>Bipolar I disorder Bipolar disorder (NOS)</td>
<td>Inpatient</td>
<td>Relapse prevention</td>
<td>Not specified</td>
<td>101</td>
<td>3</td>
<td>36</td>
<td>60</td>
<td>21</td>
<td>SADS-S</td>
<td>Psychoeducatio n</td>
<td>Group therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Faridhosseini 2017, Iran</td>
<td>Bipolar disorder</td>
<td>Outpatient</td>
<td>Quality of life</td>
<td>Not depressed</td>
<td>26</td>
<td>6</td>
<td>4</td>
<td>60</td>
<td>8</td>
<td>HDRS</td>
<td>Psychoeducatio n</td>
<td>Group therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Lam 2000, UK</td>
<td>Bipolar I disorder</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>25</td>
<td>6</td>
<td>24</td>
<td>Not stated</td>
<td>12-20</td>
<td>BDI</td>
<td>CBT</td>
<td>Individual face to face therapy</td>
<td>Usual care</td>
</tr>
<tr>
<td>Study</td>
<td>Disorder</td>
<td>Setting</td>
<td>Focus</td>
<td>Depression</td>
<td>Duration</td>
<td>Baseline</td>
<td>Follow-up</td>
<td>Comparator</td>
<td>Treatment</td>
<td>Comparison</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>------------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faria 2014, Brazil</td>
<td>Bipolar II disorder</td>
<td>Outpatient</td>
<td>Other</td>
<td>Not specified</td>
<td>61</td>
<td>N/A</td>
<td>6</td>
<td>60</td>
<td>HDRS</td>
<td>Psychoeducation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith 2011, Uk</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Quality of life</td>
<td>Not depressed</td>
<td>50</td>
<td>6</td>
<td>16</td>
<td>Not stated</td>
<td>MADRS</td>
<td>Online / e supported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morriss 2016, Uk</td>
<td>Bipolar disorder (NOS)</td>
<td>Unspecified</td>
<td>Unclear/Unspecified</td>
<td>Not depressed</td>
<td>304</td>
<td>24</td>
<td>26</td>
<td>120</td>
<td>GRID-HDRS</td>
<td>Group therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castle 2007, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not specified</td>
<td>20</td>
<td>3</td>
<td>12</td>
<td>90</td>
<td>MADRS</td>
<td>Group therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiss 2009, USA</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Other</td>
<td>Not specified</td>
<td>61</td>
<td>3</td>
<td>12</td>
<td>60</td>
<td>HDRS</td>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lam 2003, Uk</td>
<td>Bipolar I disorder</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>103</td>
<td>6</td>
<td>24</td>
<td>60</td>
<td>BDI</td>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castle 2010, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>84</td>
<td>9</td>
<td>12</td>
<td>90</td>
<td>MADRS</td>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isasi 2010, Spain</td>
<td>Bipolar I disorder</td>
<td>Unclear/Unspecified</td>
<td>Not depressed</td>
<td></td>
<td>40</td>
<td>12</td>
<td>20</td>
<td>90</td>
<td>BDI</td>
<td>CBT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardoso 2014, Brazil</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Quality of life</td>
<td>Not specified</td>
<td>61</td>
<td>6</td>
<td>6</td>
<td>60</td>
<td>HDRS</td>
<td>Individual face to face therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zaki 2014, Egypt</td>
<td>Bipolar disorder (NOS)</td>
<td>Unclear/Unspecified</td>
<td>Not specified</td>
<td></td>
<td>111</td>
<td>3</td>
<td>36</td>
<td>90-120</td>
<td>HDRS</td>
<td>Group therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ball 2006, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Relapse prevention</td>
<td>Not depressed</td>
<td>52</td>
<td>12</td>
<td>20</td>
<td>60</td>
<td>BDI</td>
<td>Individual face to face therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N/A = Not applicable
MADRS = Montgomery-Asberg Depression Rating Scale
HDRS = Hamilton Depression Rating Scale
GRID-HDRS = Global Rating of Improvement in Depression-Hamilton Rating Scale
MADRS = Montgomery-Asberg Depression Rating Scale
BDI = Beck Depression Inventory
CBT = Cognitive Behavioral Therapy
Usual Care = Usual care
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of Disorder</th>
<th>Setting</th>
<th>Type of Depression</th>
<th>Number of Individuals</th>
<th>MADRS Score</th>
<th>Comparator</th>
<th>Therapies</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miklowitz 2007, USA</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Acute depression</td>
<td>Depressed</td>
<td>293</td>
<td>Up to 30</td>
<td>CBT, family therapy, IPSRT</td>
<td>Placebo</td>
</tr>
<tr>
<td>Docteur 2020, France</td>
<td>Bipolar I disorder</td>
<td>Outpatient</td>
<td>Acute depression</td>
<td>Not specified</td>
<td>99</td>
<td>BDI-13</td>
<td>MBCT</td>
<td>Waiting list</td>
</tr>
<tr>
<td>Gliddon 2018, Australia</td>
<td>Bipolar disorder (NOS)</td>
<td>Unclear/Unspecified</td>
<td>Acute depression</td>
<td>Not specified</td>
<td>304</td>
<td>MADRS</td>
<td>Psychoeducation</td>
<td>Placebo</td>
</tr>
<tr>
<td>Jones 2019, UK</td>
<td>Bipolar disorder (NOS)</td>
<td>Unclear/Unspecified</td>
<td>Other</td>
<td>Not specified</td>
<td>44</td>
<td>HAM-D</td>
<td>CBT</td>
<td>Usual care</td>
</tr>
<tr>
<td>Lin 2020, Taiwan</td>
<td>Bipolar disorder (NOS)</td>
<td>Outpatient</td>
<td>Other</td>
<td>Not specified</td>
<td>68</td>
<td>HAMD-17</td>
<td>Psychoeducation</td>
<td>Usual care</td>
</tr>
<tr>
<td>Porter 2019, New Zealand</td>
<td>Bipolar disorder (NOS)</td>
<td>Unclear/Unspecified</td>
<td>Other</td>
<td>Not specified</td>
<td>100</td>
<td>MADRS</td>
<td>IPSRT</td>
<td>Placebo</td>
</tr>
<tr>
<td>Chen, 2018</td>
<td>Bipolar I disorder</td>
<td>Inpatient</td>
<td>Relapse prevention</td>
<td>Not specified</td>
<td>140</td>
<td>HDRS</td>
<td>Psychoeducation</td>
<td>Group</td>
</tr>
</tbody>
</table>

Note. NOS, Not otherwise specified; CBT, Cognitive Behavior Therapy; IPSRT, and Interpersonal and Social Rhythm Therapy; MADRS, Montgomery-Asberg Depression Rating Scale; BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory II; BDI 13, Beck Depression Inventory 13; HDRS, Hamilton Depression Rating Scale; GRID HDRS, GRID Hamilton Depression Rating Scale; HADS, Hospital Anxiety and Depression Scale; HAM-D, Hamilton Depression Rating Scale; HAMD-17, Hamilton Depression Rating Scale 17.
<table>
<thead>
<tr>
<th>Treatment Combination</th>
<th>Limitations</th>
<th>Heterogeneity</th>
<th>Publication Bias</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychoeducation versus usual care</td>
<td>Serious limitations</td>
<td>Substantial heterogeneity</td>
<td>No serious</td>
<td>Undetected</td>
</tr>
<tr>
<td>CBT versus usual care</td>
<td>Serious limitations</td>
<td>No evidence of significant heterogeneity</td>
<td>No serious</td>
<td>Evidence for publication bias</td>
</tr>
<tr>
<td>DBT</td>
<td>Serious limitations</td>
<td>Not applicable</td>
<td>Very serious</td>
<td>No serious</td>
</tr>
<tr>
<td>MBCT</td>
<td>Serious limitations</td>
<td>Possible moderate heterogeneity</td>
<td>Very serious</td>
<td>No serious</td>
</tr>
<tr>
<td>FT &amp; IPSRT</td>
<td>Serious limitations</td>
<td>Not applicable</td>
<td>No serious</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Psychoeducation versus active control</td>
<td>Serious limitations</td>
<td>No evidence of significant heterogeneity</td>
<td>Very serious</td>
<td>No serious</td>
</tr>
<tr>
<td>CBT versus placebo</td>
<td>Serious limitations</td>
<td>No evidence of significant heterogeneity</td>
<td>Very serious</td>
<td>No serious</td>
</tr>
<tr>
<td>CBT versus usual care</td>
<td>Serious limitations</td>
<td>No evidence of significant heterogeneity</td>
<td>No serious</td>
<td>Evidence for publication bias</td>
</tr>
<tr>
<td>MBCT versus usual care</td>
<td>Serious limitations</td>
<td>Not applicable</td>
<td>Very serious</td>
<td>No serious</td>
</tr>
<tr>
<td>MBCT versus waiting list</td>
<td>Serious limitations</td>
<td>Not applicable</td>
<td>No serious</td>
<td>No serious</td>
</tr>
</tbody>
</table>

Table 2: GRADE evidence profile
### 3.2. Study quality (Risk of Bias)

#### Figure 2. Cochrane risk of bias assessment for RCTs

<table>
<thead>
<tr>
<th>Study</th>
<th>Random sequence generation (selection bias)</th>
<th>Allocation concealment (selection bias)</th>
<th>Blinding of participants and personnel (performance bias)</th>
<th>Blinding of outcome assessment (selection bias)</th>
<th>Incomplete outcome data (attrition bias)</th>
<th>Selective reporting (reporting bias)</th>
<th>Intention to treat analysis (other bias)</th>
<th>Group similarity at baseline (other bias)</th>
<th>Manualization (other bias)</th>
<th>Training (other bias)</th>
<th>Fidelity (other bias)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D'Souza 2010</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Costa 2011</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Schmitz 2002</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Dijk 2013</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Deckerbach 2018</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Weiss 2007</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Perich 2013</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Miklowitz 2000</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Faridhosseini 2017</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Lam 2000</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Faria 2014</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Smith 2011</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Morriss 2016</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Castle 2007</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Weiss 2009</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Lam 2003</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Castle 2010</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Isasi 2010</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Cardoso 2014</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Zaki 2014</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Ball 2006</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Miklowitz 2007</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Docteur 2020</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Gliddon 2018</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Jones 2019</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Lin 2020</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Porter 2019</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
<tr>
<td>Chen 2018</td>
<td><img src="Image" alt="Random sequence generation" /></td>
<td><img src="Image" alt="Allocation concealment" /></td>
<td><img src="Image" alt="Blinding of participants" /></td>
<td><img src="Image" alt="Blinding of outcome assessment" /></td>
<td><img src="Image" alt="Incomplete outcome data" /></td>
<td><img src="Image" alt="Selective reporting" /></td>
<td><img src="Image" alt="Intention to treat analysis" /></td>
<td><img src="Image" alt="Group similarity at baseline" /></td>
<td><img src="Image" alt="Manualization" /></td>
<td><img src="Image" alt="Training" /></td>
<td><img src="Image" alt="Fidelity" /></td>
</tr>
</tbody>
</table>
Figure 2 shows the risk of bias assessment for each RCT. Three studies were at high risk of bias for random sequence generation, whereas it was not sufficiently reported in 8 of the 28 studies that were included. Only one study (Faridhosseini 2017) was at high risk of bias for allocation concealment while it was not sufficiently reported in 14 studies. Blinding of participants and personnel was not possible in 8 studies, while it was unclear for the rest. High risk of bias for detection bias was found in only one study (Weiss 2007), although detection bias was not reported sufficiently in more than half of the studies. Nineteen studies employed intention-to-treat analysis, whereas only three studies used per protocol analysis. Two studies were at high risk of bias for group similarity at baseline. Three studies were at low risk of bias for selective reporting, whereas the rest were unclear. More than two third of studies were at low risk of bias for training, although the rest were unclear. Twelve studies were at low risk of bias for fidelity, although the rest were unclear.

Overall, the quality of studies was low, other than for studies examining psychoeducation versus placebo, in which the quality was moderate.

3.3. Post-treatment effects

After extracting the data, we decided to use continuous depression score as the outcome variable in our meta-analysis as only two studies reported a categorical variable, whilst all reported a continuous variable.

3.3.1. Psychoeducation

It was possible to combine data for comparisons with usual care and placebo only. There were 1046 participants from 12 RCTs of psychoeducation who were identified through the search and included in the analysis. Psychoeducation was not found to be superior to either usual care or placebo, (Figure 3). There was a significant effect in a study that compared psychoeducation with active control (Zaki 2014; $z = 7.70$, $p < 0.00001$, SMD = -1.99 [95% CI −2.50, -1.49]).

3.3.2. CBT

There were 614 participants from 8 RCTs of CBT. Intervention was compared with usual care and placebo. Treatment modalities were group therapy and individual face-to-face therapy. There was a significant effect for comparison with usual care (figure 4). The degree of heterogeneity was low ($p = 0.31$, $I^2 = 15\%$). All outcomes from studies of CBT comparing usual care were in favour of intervention.

3.3.3. MBCT

There were 194 participants from 2 RCTs (n= 194) of MBCT, compared with usual care and waiting list. Treatment modality was group therapy. The degree of heterogeneity was moderate ($p = 0.17$, $I^2 = 47\%$). There was a significant effect in the study that compared with usual care ($z = 2.25$, $p = 0.02$, SMD = -0.47 [95% CI −0.88, -0.06]) but not in the study comparing to waiting list ($z = 0.31$, $p = 0.75$, SMD = -0.07 [95% CI −0.47, 0.34]).

3.3.4. Other therapies

There were 24 participants from 1 RCT of DBT, 49 participants from 1 RCT of family therapy, and 72 participants from 1 RCT of IPSRT who were identified through the search. Although intervention was compared with waiting list for DBT, intervention was compared with placebo for other two interventions. Treatment modality was group therapy for all.
There was a significant effect of DBT on symptoms of depression at post-treatment ($z = 2.63, p = 0.009, \text{SMD} = -1.18$ [95% CI $-2.06, -0.30$]). The effect was not significant for family therapy ($z = 0.14, p = 0.89, \text{SMD} = -0.04$ [95% CI $-0.62, 0.53$]) nor for IPSRT ($z = 0.63, p = 0.53, \text{SMD} = -0.15$ [95% CI $-0.62, 0.32$]).

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Intervention</th>
<th>Control</th>
<th>Std. Mean Difference</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>1.2.1 Psychoeducation versus usual care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardoso 2014</td>
<td>9.41</td>
<td>5.56</td>
<td>19</td>
<td>10.86</td>
</tr>
<tr>
<td>Castle 2010</td>
<td>10.6</td>
<td>7.6</td>
<td>32</td>
<td>9.4</td>
</tr>
<tr>
<td>D'Souza 2010</td>
<td>2.7</td>
<td>5.6</td>
<td>26</td>
<td>3.8</td>
</tr>
<tr>
<td>Faria 2014</td>
<td>12.53</td>
<td>8.14</td>
<td>32</td>
<td>12.62</td>
</tr>
<tr>
<td>Faridhosseini 2017</td>
<td>8.25</td>
<td>5.59</td>
<td>12</td>
<td>5.83</td>
</tr>
<tr>
<td>Lin 2020</td>
<td>7.08</td>
<td>4.7</td>
<td>38</td>
<td>9.5</td>
</tr>
<tr>
<td>Miklowitz 2000</td>
<td>2</td>
<td>0.8</td>
<td>28</td>
<td>2.4</td>
</tr>
<tr>
<td>Smith 2011</td>
<td>9.1</td>
<td>8.4</td>
<td>17</td>
<td>11.1</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>204</td>
<td>235</td>
<td>100.0%</td>
<td>-0.15 [-0.34, 0.05]</td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 7.14$, df = 7 ($p = 0.41$); $I^2 = 2$
Test for overall effect: $Z = 1.46$ ($P = 0.14$)

1.2.3 Psychoeducation versus placebo

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Intervention</th>
<th>Control</th>
<th>Std. Mean Difference</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>Castle 2007</td>
<td>13</td>
<td>7.3</td>
<td>8</td>
<td>5.8</td>
</tr>
<tr>
<td>Gilidon 2019</td>
<td>9.9</td>
<td>7.7</td>
<td>202</td>
<td>12</td>
</tr>
<tr>
<td>Moriss 2016</td>
<td>7</td>
<td>7.17</td>
<td>122</td>
<td>8.18</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>332</td>
<td>176</td>
<td>100.0%</td>
<td>-0.10 [-0.45, 0.25]</td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.05$; $\chi^2 = 4.53$, df = 2 ($p = 0.10$); $I^2 = 56$
Test for overall effect: $Z = 0.57$ ($P = 0.57$)

Test for subgroup differences: $\chi^2 = 0.05$, df = 1 ($P = 0.82$), $I^2 = 0$

Figure 3. Forest plot of post-treatment depression outcome for Psychoeducation versus different comparators.
3.4. Follow up effects
3.4.1. Psychoeducation (3 to 12 months)

There were 626 participants from 6 RCTs of psychoeducation who were included in the analysis. Psychoeducation for bipolar disorder had no effect on symptoms of depression at follow up among 4 studies comparing it to usual care and another two studies comparing it to placebo.

3.4.2. CBT (3 to 12 months)

There were 482 participants from 6 RCTs of CBT who were included in the analysis. One study (Isasi 2010) had significantly better depression outcome than did the others. The degree of heterogeneity was high (p = 0.0001, I² = 80%). CBT for bipolar disorder had no effect on symptoms of depression at follow up among studies comparing it to usual care nor in another study comparing it to placebo (Gliddon 2018; z = 1.99, p = 0.05, SMD = -0.32 [95% CI −0.64, -0.00]).

3.4.3. MBCT (12 months)

There were 95 participants from 1 RCTs of MBCT. MBCT for bipolar disorder had no effect on symptoms of depression at follow up (z = 0.90, p = 0.37, SMD = 0.18 [95% CI −0.22, 0.59]).
3.5. Publication bias

There was no evidence of publication bias detected overall according to Egger's test (P = 0.493). However, when subgroups of studies were examined based on therapy type, there was an evidence for publication bias for studies of CBT.

3.6. Sensitivity Analysis

In this meta-analysis, we aimed to conduct sensitivity analysis to examine the effect of (i) intervention target, (ii) pre-treatment depression-status; (iii) modality of therapy (self-help,
individual face-to-face therapy, group therapy), (vi) group versus individual therapy, (v) psychological comorbidity at baseline.

As only one study had presence of acute depression as an inclusion criterion it was not possible to run a sensitivity analysis based on this variable. The same applied to the analysis pertaining to presence of psychological comorbidity at baseline.

3.6.1. Intervention target

The target of the therapy was clear in 14 of 28 studies. Of these, 3 targeted acute depression, 8 targeted relapse prevention and 3 quality of life. Within these subgroups, there was variability in both therapy type (e.g. CBT, psychoeducation) and comparator (e.g. usual care, placebo). Figure 7 shows the effect sizes for therapies targeting acute depression versus those targeting quality of life and relapse prevention. Given the small number of studies, we were not able to conduct meaningful direct comparison between those targeting acute depression and those targeting other issues. Examining the effect sizes for acute depression trials alone, we note that despite heterogeneity in therapy type and in comparator, the effect sizes are tightly clustered around a small overall effect size (SMD = -0.119 [95% CI −0.280, 0.043]).

3.6.2. Pre-treatment depression-status

Pre-treatment depression status was operationalised in two ways: i) whether or not the study inclusion criteria required participants to be depressed; ii) mean depression score for the study sample pre-treatment. As previously stated, analysis according to (i) was not possible because only one study required participants to be acutely depressed. With regard to (ii), the number of studies was too small to allow meta-regression within subgroups according to treatment type and comparator. Instead, we conducted meta-regression on the full set of studies, controlling for the study characteristics of therapy type and comparator. The result showed no significant relationship between baseline depression score and depression outcome post-treatment, when controlling for therapy type and comparator (p = 0.829, 0.003 [-0.026, 0.032]).

3.6.3. Group versus individual therapy

Additional sensitivity analysis showed that there was no significant difference in depression outcome between studies using individual (14 studies; SMD = -0.322 [95% CI −0.529, −0.116]) versus group therapy (6 studies; SMD = -0.352 [95% CI −0.660, −0.044]).
4. Discussion

The purpose of this review was to evaluate the effectiveness of psychological interventions for patients with bipolar I or II disorder in reducing the symptoms of depression. We also set out to examine whether intervention target, pre-treatment depression-status, modality of therapy, group versus individual therapy, and psychological comorbidity at baseline impact upon treatment outcomes.

Considering first the evidence for the impact of therapy on depression post-treatment, for CBT and DBT there was low quality evidence of a significant effect on depression post-treatment, particularly in comparison with usual care. At long term follow-up, there was no effect on depression.

Whilst one meta-analysis (Bi-Yu et al., 2016) found no impact of CBT on depression symptoms, our findings support the more recent meta-analyses (Miklowitz, 2021; Chiang et al., 2017) in finding that CBT has benefit upon depressive symptoms post-treatment. This discrepancy may be due to differences in the studies included in the respective meta-
analyses. We note that study numbers in subgroup comparisons in our and other meta-
analyses are small, and in our meta-analysis only one study was included that tested DBT.
We also note that meta-analyses differed in their definitions around the time point for post-
treatment and follow-up outcomes. In this study, we defined post-treatment outcome as
the point immediately following the end of the acute treatment phase, and no later than 3
months after the end of treatment. However, this may differ both from other meta-analyses
and from the primary outcome points reported in the original study papers. We consider
this a strength of the current study in that we are examining depression outcome
immediately following treatment, and this is consistent across included studies.
Psychoeducation was not significantly better than usual care and placebo. However,
studies comparing psychoeducation to active control found a benefit. We found no
significant effects for MBCT, family therapy, and IPSRT at post-treatment. There were only a
small number of studies in these analyses, and no data on longer term follow-up.
When covarying for therapy type and comparator, we found no significant link between
baseline depression score and depression outcome post-treatment. More studies using
similar therapy types and comparators are needed to be able to better look at the
relationship between depression status at baseline and outcome.
Given the small number of studies, we were not able to conduct meaningful direct
comparison between those targeting acute depression and those targeting other issues.
Nevertheless our forest plot revealed the effect sizes for studies targeting acute depression
to be tightly clustered around a small overall effect size (SMD = -0.12). This is smaller than
the comparable effect size found for unipolar depression, for studies comparing CBT to
usual care (SMD = 0.59) (Cuijpers et al., 2013), although it is difficult to make a direct
comparison due to heterogeneity amongst the acute depression studies in our analysis in
terms of therapy modality and comparator. Considering the effect size for depression
outcome we obtained for CBT compared to usual care regardless of therapy target (SMD = - 0.51) we note that this is similar to the effect size obtained by Cuijpers and colleagues
(2013), although some of the trials in our analysis focussed on relapse, rather than acute
depression treatment.
Across all therapy modalities, sensitivity analysis found no significant difference in post-
treatment depression outcome between studies using individual and group therapy.
Because of limited number of studies it was not possible to separately examine studies in
outpatient and inpatient settings.

4.1. Implications
More high-quality randomised controlled trials of psychological interventions for bipolar
disorder are required. We would also recommend that studies clearly identify the primary
difficulty or difficulties the therapy protocol seeks to target (for example, acute depression,
relapse prevention, anxiety). Furthermore, it appears that further development and testing
of approaches targeting acute bipolar depression is required.
In terms of reducing depression post-treatment, based upon our analysis the evidence
favours CBT, regardless of whether relapse prevention or acute depression is the target,
however from the studies we included we cannot determine whether CBT performs better
than other therapies, as there were few direct comparisons. This leaves open the question
of whether there could be further increases in the efficacy of CBT for bipolar depression if
acute depression is the primary focus in the protocol: more studies are needed to address
this question. Furthermore, recommended interventions for bipolar depression are
largely based on psychological models of unipolar depression: it is possible they may be more effective if derived from evidence-based models of bipolar depression.

4.2. Limitations

Our study is subject to some limitations. First, we may not have included all relevant studies as we excluded non-English-language papers and therefore might have missed some relevant trials published in other languages. In addition, some potentially relevant studies were excluded because means and standard deviations of depression scores were not reported in the published paper, and data were unavailable from the authors.

Second, we were limited in our ability to conduct the planned sensitivity analyses. We could not perform an overall analysis of the effect of treatment target because of heterogeneity amongst a small set of studies. Because there were too few separate trials that explicitly selected depressed participants, we were not able to run a sensitivity analysis of depression status at baseline, as measured by stated study inclusion criteria. Also, our meta-regression looking at the impact of baseline depression was conducted upon a relatively small set of studies with additional degrees of freedom lost due to the need to include covariates. We were not able to look at the impact of treatment modality (remote versus face to face) because too few studies were delivered remotely. In addition, as only one study had presence of acute depression and psychological comorbid condition at baseline as inclusion criteria, we were not able to run a sensitivity analysis using these covariates.

Third, we recognise that our definition of psychological interventions (non-pharmacological treatments whereby the primary hypothesized or intended mechanism of action is via changes in cognitive, behavioural, relational, psychodynamic or interpersonal processes) may exclude some non-medical, relationship-based treatments such as collaborative care.

5. Conclusion

Some psychological therapies such as CBT and DBT appear to reduce acute bipolar depression post-treatment, but this effect is not present at longer term follow-up. In the absence of a greater number of high-quality studies however, this conclusion should be viewed with caution. In a small set of studies, the relationship between baseline depression score and depression outcome post-treatment was found not to be significant: there is a need for more research employing similar therapy types and comparators, and defining depression status at baseline, in order to better understand the association between depression status at baseline and outcome. As a result of the small number of studies, we could not directly compare those targeting acute depression to those targeting other issues, nevertheless studies targeting acute depression showed small overall effect sizes that are equivalent or smaller than the equivalent reported for unipolar depression. Further development and testing of interventions specifically targeting bipolar depression is required.

Funding

This study is partly supported by Ministry of National Education of Turkey.

Declarations of interest

The authors declare no conflicts of interest.
References


