How effective are cognitive behavior therapies for major depression and anxiety disorders?

Citation for published version (APA):


Document status and date:
Published: 01/10/2016

DOI: 10.1002/wps.20346

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at: repository@maastrichtuniversity.nl
providing details and we will investigate your claim.

Download date: 12 Jul. 2023
How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence

Pim Cuijpers1,2, Ioana A. Cristea3, Erini Karyotaki1,2, Mirjam Reijnders1,2, Marcus J.H. Huibers1,2,4

1Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; 2EMGO Institute for Health and Care Research, Amsterdam, The Netherlands; 3Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Cluj-Napoca, Romania; 4Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

We report the current best estimate of the effects of cognitive behavior therapy (CBT) in the treatment of major depression (MDD), generalized anxiety disorder (GAD), panic disorder (PAD) and social anxiety disorder (SAD), taking into account publication bias, the quality of trials, and the influence of waiting list control groups on the outcomes. In our meta-analyses, we included randomized trials comparing CBT with a control condition (waiting list, care-as-usual or pill placebo) in the acute treatment of MDD, GAD, PAD or SAD, diagnosed on the basis of a structured interview. We found that the overall effects in the 144 included trials (184 comparisons) for all four disorders were large, ranging from $g=0.75$ for MDD to $g=0.80$ for GAD, $g=0.81$ for PAD, and $g=0.88$ for SAD. Publication bias mostly affected the outcomes of CBT in GAD (adjusted $g=0.59$) and MDD (adjusted $g=0.65$), but not those in PAD and SAD. Only 17.4% of the included trials were considered to be high-quality, and this mostly affected the outcomes for PAD ($g=0.61$) and SAD ($g=0.76$). More than 80% of trials in anxiety disorders used waiting list control groups, and the few studies using other control groups pointed at much smaller effect sizes for CBT. We conclude that CBT is probably effective in the treatment of MDD, GAD, PAD and SAD; that the effects are large when the control condition is waiting list, but small to moderate when it is care-as-usual or pill placebo; and that, because of the small number of high-quality trials, these effects are still uncertain and should be considered with caution.

Key words: Cognitive behavior therapy, major depression, generalized anxiety disorder, panic disorder, social anxiety disorder, meta-analysis, publication bias, quality of trials, waiting list control groups

Every year almost 20% of the general population suffers from a common mental disorder, such as depression or an anxiety disorder. These conditions not only result in personal suffering for patients and their families, but also in huge economic costs, in terms of both work productivity loss and health and social care expenditures. Several evidence-based treatments are available for common mental disorders, including pharmacological and psychological interventions. Many patients receive pharmacological treatments, and these numbers are increasing in high-income countries. Psychological treatments are equally effective in the treatment of depression and anxiety disorders. However, they are less available or accessible, especially in low- and middle-income countries. At the same time, about 75% of patients prefer psychotherapy over the use of medication.

The most extensively tested form of psychotherapy is cognitive behavior therapy (CBT). Dozens of trials and several meta-analyses have shown that CBT is effective in treating depression and anxiety disorders. However, in recent years, it has become clear that the effects of CBT and other psychotherapies have been considerably overestimated due to at least three reasons.

The first reason is publication bias. This refers to the tendency of authors to submit, or journals to accept, manuscripts for publication based on the direction or strength of the study’s findings. There is considerable indirect evidence of publication bias in psychotherapy research, based on excess publication of small studies with large effect sizes. Moreover, there is also direct evidence of publication bias: a recent study found that almost one quarter of trials of psychotherapy for adult depression funded by the US National Institutes of Health were not published. After adding the effect sizes of these unpublished trials to those of the published ones, the mean effect size for psychotherapy dropped by more than 25%.

The second reason why the effects of psychotherapies have been overestimated is that the quality of many trials is suboptimal. In a meta-analysis of 115 trials of psychotherapy for depression, only 11 met all basic indicators of quality, and the effect sizes of these trials were considerably smaller than those of lower quality ones. However, that meta-analysis only included trials up to 2008, and since then many new studies have been conducted. Because more recent trials are typically of a better quality than older ones, it is not known what the current best estimate of the effect size of CBT is after taking these newer studies into account.

A third reason why the effects of psychotherapy have been overestimated is that many trials have used waiting list control groups. Although all control conditions in psychotherapy trials have their own problems, the improvement found in patients on waiting lists has been found to be lower than that expected on the basis of spontaneous remission. It has been suggested, therefore, that waiting list is in fact a “nocebo” (the opposite of a placebo; an inert treatment that appears to cause an adverse effect) and that trials using it considerably overestimate the effects of psychological treatments. Other control conditions, such as care-as-usual and pill placebo, can allow a better estimate of the true effect size of CBT.

In the present paper, we report the most up-to-date and accurate estimate of the effects of CBT in the treatment of
major depression (MDD), generalized anxiety disorder (GAD), panic disorder (PAD) and social anxiety disorder (SAD), taking into account the three above-mentioned major problems of the existing psychotherapy research: publication bias, low quality of trials, and the nocebo effect of waiting list control groups.

METHODS

Identification and selection of studies

We searched four major bibliographic databases (PubMed, PsycINFO, Embase and the Cochrane database of randomized trials) by combining terms (both MeSH terms and text words) indicative of psychological treatment and either SAD (social phobia, social anxiety, public-speaking anxiety), GAD (worry, generalized anxiety), or PAD with or without agoraphobia (panic, panic disorder), with filters for randomized controlled trials. We also checked the references of earlier meta-analyses on psychological treatments for the included disorders. The deadline for the searches was August 14, 2015.

For the identification of trials of CBT for MDD, we used an existing database22 updated to January 2016 by combining terms indicative of psychological treatment and depression (both MeSH terms and text words).

We included randomized trials in which CBT was directly compared with a control condition (waiting list, care-as-usual or pill placebo) in adults with MDD, GAD, PAD or SAD. Only trials in which recruited subjects met diagnostic criteria for the disorder according to a structured diagnostic interview – such as the Structured Clinical Interview for DSM (SCID), the Composite International Diagnostic Interview (CIDI) or the Mini International Neuropsychiatric Interview (MINI) – were included.

In addition to any therapy in which cognitive restructuring was one of the core components, we also included purely behavioral therapies, i.e., trials of behavioral activation for depression and exposure for anxiety disorders. We included therapies that used individual, group and guided self-help formats, but excluded self-guided therapies without any professional support, because their effects have been found to be considerably smaller than other formats23. Studies on therapies delivering only (applied) relaxation were excluded, as were studies on eye movement desensitization and reprocessing (EMDR), interpersonal or psychodynamic therapy, virtual reality therapy, transdiagnostic therapies, as well as studies in which CBT was combined with pill placebo.

In order to keep heterogeneity as low as possible, we included only studies using waiting list, care-as-usual or pill placebo control groups. Care-as-usual was defined broadly as anything patients would normally receive, as long as it was not a structured type of psychotherapy. Psychological placebo conditions were not included, because they have considerable effects on depression24 and probably also on anxiety disorders19. Comorbid mental or somatic disorders were not used as an exclusion criterion. Studies on inpatients and on adolescents or children (below 18 years of age) were excluded, as were studies recruiting patients with other types of depressive disorders than MDD (dysthymia or minor depression). We also excluded maintenance studies, aimed at people who already had a partial or complete remission after an earlier treatment, and studies that did not report sufficient data to calculate standardized effect sizes. Studies in English, German and Dutch were considered for inclusion.

Quality assessment and data extraction

We assessed the quality of included studies using four criteria of the “risk of bias” assessment tool developed by the Cochrane Collaboration25. Although “risk of bias” and quality are not synonyms25, the former can be seen as an indicator of the quality of studies. The four criteria were: adequate generation of allocation sequence; concealment of allocation to conditions; blinding of assessors; and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat analyses were conducted, meaning that all randomized patients were included in the analyses). The assessment of the quality of included studies was conducted by two independent researchers, and disagreements were solved through discussion.

We also coded participant characteristics (disorder, recruitment method, target group); characteristics of the psychotherapies (treatment format, number of sessions); and general characteristics of the studies (country where the study was conducted, year of publication).

Meta-analyses

For each comparison between a psychotherapy and a control condition, the effect size indicating the difference between the two groups at post-test was calculated (Hedges’ g). Effect sizes of 0.8 can be assumed to be large, while effect sizes of 0.5 are moderate, and effect sizes of 0.2 are small26. Effect sizes were determined by subtracting (at post-test) the average score of the psychotherapy group from the average score of the control group, and dividing the result by the pooled standard deviation. Because some studies had relatively small sample sizes, we corrected the effect size for small sample bias27. If means and standard deviations were not reported, we calculated the effect size using dichotomous outcomes, and if these were not available either, we used other statistics (such a t or p value).

In order to calculate effect sizes, we used all measures examining depressive symptoms, such as the Beck Depression Inventory (BDI28 or BDI-II29) and the Hamilton Rating Scale for Depression30, or anxiety symptoms, such as the Beck Anxiety Inventory31, the Penn State Worry Questionnaire32, the Fear Questionnaire33, and the Liebowitz Social Anxiety Scale34. We did not use measures of mediators, dysfunctional thinking, quality of life or generic severity. To calculate pooled mean effect sizes, we used the Comprehensive Meta-Analysis (CMA)
software (version 3.3.070). Because we expected considerable heterogeneity among the studies, we employed a random effects pooling model in all analyses.

Numbers-needed-to-treat (NNT) were calculated using the formulae provided by Furukawa35, in which the control group's event rate was set at a conservative 19% (based on the pooled response rate of 50% reduction of symptoms across trials of psychotherapy for depression)36. As a test of homogeneity of effect sizes, we calculated the I² statistic (a value of 0 indicates no observed heterogeneity, and larger values indicate increasing heterogeneity, with 25 as low, 50 as moderate, and 75 as high heterogeneity)37. We calculated 95% confidence intervals around I² using the non-central chi-squared-based approach within the Heterogi module for Stata38,39.

We conducted subgroup analyses according to the mixed effects model, in which studies within subgroups are pooled with the random effects model, while tests for significant differences between subgroups are conducted with the fixed effects model. For continuous variables, we used meta-regression analyses to test whether there was a significant relationship between the continuous variable and the effect size, as indicated by a Z value and an associated p value. Multivariate meta-regression analyses, with the effect size as the dependent variable, were conducted using CMA.

We tested for publication bias by inspecting the funnel plot on primary outcome measures and by Duval and Tweedie’s trim and fill procedure40, which yields an estimate of the effect size after the publication bias has been taken into account. We also conducted Egger’s test of the intercept to quantify the bias captured by the funnel plot and to test whether it was significant.

### RESULTS

#### Selection and inclusion of trials

After examining a total of 26,775 abstracts (19,580 after removal of duplicates), we retrieved 2,957 full-text papers for further consideration. We excluded 2,813 of the retrieved papers. The PRISMA flow chart describing the inclusion process and the reasons for exclusion is presented in Figure 1. A total of 144 trials met inclusion criteria for this meta-analysis: 54 on MDD, 24 on GAD, 30 on PAD, and 36 on SAD.

#### Characteristics of included trials

The 144 trials included a total of 184 comparisons between CBT and a control condition (63 comparisons for MDD, 31 for GAD, 42 for PAD, and 48 for SAD). A total of 11,030 patients were enrolled (6,229 in the CBT groups, 2,469 in the waiting list control groups, 1,823 in the care-as-usual groups and 509 in the pill placebo groups). A total of 113 trials were aimed at adults in general and 31 at other more specific target groups. Eighty trials recruited patients (also) from the community, 51 recruited exclusively from clinical populations, and 13 used other recruitment methods. Sixty-seven trials were conducted in North America, 14 in the UK, 36 in other European countries, 15 in Australia, 4 in East Asia, and 8 in other geographic areas. Of all included trials, 44 (30.6%) were conducted in 2010 or later.

CBT was delivered in individual format in 87 comparisons, in group format in 53, in guided self-help format in 35, and in a mixed or another format in 9. The number of treatment sessions ranged from one to 25.

#### Quality assessment

Sixty trials reported an adequate sequence generation, while the other 84 did not. A total of 46 trials reported allocation to conditions by an independent (third) party. Seventy trials reported blinding of outcome assessors and 57 conducted intention-to-treat analyses. Only 25 trials (17.4%) met all four quality criteria, 62 met two or three criteria, and the remaining 57 met one or none of the criteria. Of the trials conducted in 2010 or later, 29.5% were rated as high-quality, compared to 12.0% of the older studies.

#### Effects of CBT on MDD

The pooled effect size of the 63 comparisons between CBT and control conditions in MDD41-94 was $g=0.75$ (95% CI: 0.64-0.87), with high heterogeneity ($I^2=71$). This effect size corresponds to a NNT of 3.86. Studies using a waiting list control group had significantly ($p=0.002$) larger effect sizes ($g=0.98$; 95% CI: 0.80-1.17) than those using care-as-usual ($g=0.60$; 95% CI: 0.45-0.75) and pill placebo control groups ($g=0.55$; 95% CI: 0.28-0.81) (Table 1 and Figure 2).

---

**Figure 1** Flow chart of inclusion of trials. MDD – major depression, GAD – generalized anxiety disorder, PAD – panic disorder, SAD – social anxiety disorder, CBT – cognitive behavior therapy.
Only 11 of the 63 studies were rated as being high-quality. The effect size in these studies was similar to that in the total pool ($g = 0.73$; 95% CI: 0.46-1.00; $I^2 = 78$). No high-quality study used a pill placebo control group. The difference between waiting list and care-as-usual among the high-quality studies was not significant ($p = 0.06$), but this may be related to the small number of those studies.

Egger’s test indicated considerable asymmetry of the funnel plot (intercept: 1.54; 95% CI: 0.59-2.50; $p = 0.001$). Duval and Tweedie’s trim and fill procedure also indicated considerable publication bias (number of imputed studies: 8; adjusted effect size: $g = 0.65$; 95% CI: 0.53-0.78; $I^2 = 76$). For high-quality studies, no indication for publication bias was found (but this may again be related to the small number of those studies).

### Table 1: Effects of cognitive behavior therapy for major depression (MDD), generalized anxiety disorder (GAD), panic disorder (PAD) and social anxiety disorder (SAD) compared to control conditions

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>g</th>
<th>95% CI</th>
<th>p</th>
<th>$I^2$</th>
<th>95% CI</th>
<th>p</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MDD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All control conditions</td>
<td>All studies</td>
<td>63</td>
<td>0.75</td>
<td>0.64-0.87</td>
<td>&lt;0.001</td>
<td>71</td>
<td>62-77</td>
<td>3.86</td>
</tr>
<tr>
<td></td>
<td>High-quality studies</td>
<td>11</td>
<td>0.73</td>
<td>0.46-1.00</td>
<td>&lt;0.001</td>
<td>78</td>
<td>56-86</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>Adjusted for publication bias</td>
<td>71</td>
<td>0.65</td>
<td>0.53-0.78</td>
<td>&lt;0.001</td>
<td>76</td>
<td>69-80</td>
<td>4.55</td>
</tr>
<tr>
<td>Type of control</td>
<td>Waiting list</td>
<td>28</td>
<td>0.98</td>
<td>0.80-1.17</td>
<td>&lt;0.001</td>
<td>68</td>
<td>50-77</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>30</td>
<td>0.60</td>
<td>0.45-0.75</td>
<td>&lt;0.001</td>
<td>69</td>
<td>54-78</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>Pill placebo</td>
<td>5</td>
<td>0.55</td>
<td>0.28-0.81</td>
<td>&lt;0.001</td>
<td>45</td>
<td>0-78</td>
<td>5.51</td>
</tr>
<tr>
<td>High-quality studies</td>
<td>Waiting list</td>
<td>6</td>
<td>0.93</td>
<td>0.49-1.57</td>
<td>&lt;0.001</td>
<td>82</td>
<td>56-90</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>5</td>
<td>0.43</td>
<td>0.16-0.70</td>
<td>&lt;0.001</td>
<td>46</td>
<td>0-79</td>
<td>7.29</td>
</tr>
<tr>
<td><strong>GAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All control conditions</td>
<td>All studies</td>
<td>31</td>
<td>0.80</td>
<td>0.67-0.93</td>
<td>&lt;0.001</td>
<td>33</td>
<td>0-56</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td>High-quality studies</td>
<td>9</td>
<td>0.82</td>
<td>0.60-1.04</td>
<td>&lt;0.001</td>
<td>46</td>
<td>0-73</td>
<td>3.49</td>
</tr>
<tr>
<td></td>
<td>Adjusted for publication bias</td>
<td>42</td>
<td>0.59</td>
<td>0.44-0.75</td>
<td>&lt;0.001</td>
<td>62</td>
<td>44-72</td>
<td>5.08</td>
</tr>
<tr>
<td>Type of control</td>
<td>Waiting list</td>
<td>24</td>
<td>0.85</td>
<td>0.72-0.99</td>
<td>&lt;0.001</td>
<td>13</td>
<td>0-47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>4</td>
<td>0.45</td>
<td>0.26-0.64</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0-68</td>
<td>6.93</td>
</tr>
<tr>
<td></td>
<td>Pill placebo</td>
<td>3</td>
<td>1.32</td>
<td>0.83-1.81</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0-73</td>
<td>2.08</td>
</tr>
<tr>
<td>High-quality studies</td>
<td>Waiting list</td>
<td>8</td>
<td>0.88</td>
<td>0.67-1.10</td>
<td>&lt;0.001</td>
<td>33</td>
<td>0-69</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>1</td>
<td>0.45</td>
<td>0.08-0.83</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0</td>
<td>6.93</td>
</tr>
<tr>
<td><strong>PAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All control conditions</td>
<td>All studies</td>
<td>42</td>
<td>0.81</td>
<td>0.59-1.04</td>
<td>&lt;0.001</td>
<td>77</td>
<td>69-82</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td>High-quality studies</td>
<td>4</td>
<td>0.61</td>
<td>0.27-0.96</td>
<td>&lt;0.001</td>
<td>26</td>
<td>0-75</td>
<td>4.89</td>
</tr>
<tr>
<td>Type of control</td>
<td>Waiting list</td>
<td>33</td>
<td>0.96</td>
<td>0.70-1.23</td>
<td>&lt;0.001</td>
<td>77</td>
<td>67-82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>4</td>
<td>0.27</td>
<td>−0.12 to 0.65</td>
<td>&lt;0.001</td>
<td>31</td>
<td>0-77</td>
<td>12.25</td>
</tr>
<tr>
<td></td>
<td>Pill placebo</td>
<td>5</td>
<td>0.28</td>
<td>0.03-0.54</td>
<td>&lt;0.001</td>
<td>8</td>
<td>0-67</td>
<td>11.77</td>
</tr>
<tr>
<td>High-quality studies</td>
<td>Waiting list</td>
<td>4</td>
<td>0.61</td>
<td>0.27-0.96</td>
<td>&lt;0.001</td>
<td>26</td>
<td>0-75</td>
<td>4.89</td>
</tr>
<tr>
<td><strong>SAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All control conditions</td>
<td>All studies</td>
<td>48</td>
<td>0.88</td>
<td>0.74-1.03</td>
<td>&lt;0.001</td>
<td>64</td>
<td>50-73</td>
<td>3.22</td>
</tr>
<tr>
<td></td>
<td>High-quality studies</td>
<td>8</td>
<td>0.76</td>
<td>0.47-1.06</td>
<td>&lt;0.001</td>
<td>71</td>
<td>25-84</td>
<td>3.80</td>
</tr>
<tr>
<td>Type of control</td>
<td>Waiting list</td>
<td>40</td>
<td>0.98</td>
<td>0.83-1.14</td>
<td>&lt;0.001</td>
<td>64</td>
<td>47-73</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Care-as-usual</td>
<td>3</td>
<td>0.44</td>
<td>0.12-0.77</td>
<td>&lt;0.001</td>
<td>23</td>
<td>0-79</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>Pill placebo</td>
<td>5</td>
<td>0.47</td>
<td>0.24-0.70</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0-64</td>
<td>6.59</td>
</tr>
<tr>
<td>High-quality studies</td>
<td>Waiting list</td>
<td>5</td>
<td>1.00</td>
<td>0.61-1.40</td>
<td>&lt;0.001</td>
<td>71</td>
<td>0-87</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Care as usual</td>
<td>2</td>
<td>0.30</td>
<td>−0.04 to 0.64</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0</td>
<td>10.91</td>
</tr>
<tr>
<td></td>
<td>Pill placebo</td>
<td>1</td>
<td>0.57</td>
<td>0.20-0.93</td>
<td>&lt;0.001</td>
<td>0</td>
<td>0</td>
<td>5.29</td>
</tr>
</tbody>
</table>

NNT – Number-needed-to-treat
Figure 2 Effects of cognitive behavior therapy (CBT) for major depression compared to control conditions: forest plot.

BA – behavioral activation, CT – cognitive therapy, in-p – in person, tel – telephone, MBCT – mindfulness based CBT, RCBT – religious CBT, SCBT – secular CBT, iCBT – Internet-delivered CBT, techn – supported by a technician, clin – supported by a clinician, e-mail – supervised by e-mail, gsh – guided self-help format, BIB – bibliotherapy, COP – coping, cCBT – computerized CBT.
The vast majority of studies (24 of 31) used a waiting list control group, so the effects of care-as-usual control groups (N = 3) and care-as-usual control groups (N = 4) was very small, however (Table 1 and Figure 3).

Only 9 of the 31 studies were rated as high-quality, and 8 of these used a waiting list control group, so the effects of care-as-usual and pill placebo among high-quality studies could not be estimated.

Egger’s test was significant (intercept: 1.60; 95% CI: 0.38-2.83; p = 0.006). Duval and Tweedie’s trim and fill procedure resulted in an adjusted effect size of g = 0.59 (95% CI: 0.44-0.75; I² = 62; number of imputed studies: 11). For high-quality studies, no indication for publication bias was found (but this may again be related to the small number of those studies).

The 42 comparisons between CBT and control conditions in PAD138-147 resulted in a pooled effect size of g = 0.81 (95% CI: 0.59-1.04; F = 77; NNT = 3.53). In the vast majority of the comparisons (N = 33), a waiting list control condition was used. The difference between studies using a waiting list (g = 0.96) and either care-as-usual (g = 0.27) or pill placebo (g = 0.28) was significant (p < 0.001). The four comparisons of CBT versus care-as-usual even indicated a non-significant effect size (g = 0.27; 95% CI: −0.12 to 0.65; p = 0.17) (Table 1 and Figure 4).

The four high-quality studies all used a waiting list control group and resulted in an effect size of g = 0.61 (95% CI: 0.27-0.96).

Although Egger’s test indicated significant asymmetry of the funnel plot (intercept: 3.62; 95% CI: 0.90-6.34; p = 0.005), Duval and Tweedie’s trim and fill procedure did not indicate any missing studies and therefore the adjusted and unadjusted effect sizes were the same. In the four high-quality studies, no indication for publication bias was found.
The 48 comparisons between CBT and a control condition resulted in a pooled effect size of $g = 0.88$ (95% CI: 0.74-1.03; $I^2 = 64$; NNT = 3.22). Again, the large majority of studies used a waiting list control group (N = 40), with only three using care-as-usual and five pill placebo. The studies using a waiting list control group resulted in significantly ($p < 0.001$) larger effect sizes ($g = 0.98$) than those using a pill placebo ($g = 0.47$) or care-as-usual control group ($g = 0.44$) (Table 1 and Figure 5).

Only eight studies were rated as high-quality, and five of these used a waiting list control group. This implies that for SAD there are not enough high-quality studies to assess the effects of CBT compared to care-as-usual or pill placebo.

Egger’s test pointed at significant asymmetry of the funnel plot (intercept: 2.46; 95% CI: 0.96-3.96; $p = 0.001$), but Duval and Tweedie’s trim and fill procedure did not indicate missing studies and the adjusted and unadjusted effect sizes were the same.

**Multivariate meta-regression analyses**

We conducted four separate analyses, for each disorder, with the effect size as the dependent variable and characteristics of

<table>
<thead>
<tr>
<th>Study</th>
<th>$g$</th>
<th>95% CI</th>
<th>p</th>
<th>$g$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addis et al</td>
<td>-0.12</td>
<td>-0.56 to 0.32</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Bakker et al</td>
<td>-0.10</td>
<td>-0.58 to 0.38</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Barlow et al</td>
<td>0.42</td>
<td>-0.33 to 1.17</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Barlow et al</td>
<td>0.49</td>
<td>-0.25 to 1.23</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Barlow et al</td>
<td>0.42</td>
<td>-0.04 to 0.88</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Black et al</td>
<td>0.18</td>
<td>-0.39 to 0.74</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Botella et al</td>
<td>1.53</td>
<td>0.64-2.43</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Carlbring et al</td>
<td>0.54</td>
<td>-0.07 to 1.15</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Carling et al</td>
<td>0.54</td>
<td>0.33-1.40</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Carter et al</td>
<td>1.64</td>
<td>0.72-2.55</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Case et al</td>
<td>2.20</td>
<td>1.55-2.85</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Clark et al</td>
<td>2.72</td>
<td>1.76-3.68</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Clark et al</td>
<td>1.73</td>
<td>0.85-2.60</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Clark et al</td>
<td>2.05</td>
<td>1.13-2.97</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Gould et al</td>
<td>0.22</td>
<td>-0.64 to 1.07</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Gould et al</td>
<td>-0.46</td>
<td>-1.34 to 0.42</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Gould et al</td>
<td>-0.37</td>
<td>-1.17 to 0.42</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Hazen et al</td>
<td>0.78</td>
<td>0.22-1.34</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Hazen et al</td>
<td>1.14</td>
<td>0.55-1.72</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Hendriks et al</td>
<td>0.98</td>
<td>0.23-1.73</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Ito et al</td>
<td>2.07</td>
<td>1.24-2.90</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Ito et al</td>
<td>2.01</td>
<td>1.20-2.82</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Ito et al</td>
<td>2.05</td>
<td>1.24-2.86</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Klein &amp; Richards</td>
<td>0.71</td>
<td>-0.14 to 1.56</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Klosko et al</td>
<td>0.63</td>
<td>-0.17 to 1.43</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Lessard et al</td>
<td>0.46</td>
<td>-0.25 to 1.17</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Lessard et al</td>
<td>0.43</td>
<td>-0.24 to 1.10</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Lidren et al</td>
<td>0.61</td>
<td>-0.26 to 1.48</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Lidren et al</td>
<td>0.05</td>
<td>-0.79 to 0.89</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>Ross &amp; Scott</td>
<td>-0.07</td>
<td>-0.88 to 0.73</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Schmidt et al</td>
<td>1.60</td>
<td>0.75-2.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Schmidt et al</td>
<td>1.60</td>
<td>0.73-2.47</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Schmidt et al</td>
<td>1.94</td>
<td>1.24-2.65</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Sharp et al</td>
<td>0.51</td>
<td>-0.04 to 1.06</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Sharp et al</td>
<td>0.25</td>
<td>-0.38 to 0.89</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Sharp et al</td>
<td>0.61</td>
<td>0.00-1.22</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Swinson et al</td>
<td>1.01</td>
<td>0.37-1.65</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Telch et al</td>
<td>0.93</td>
<td>0.43-1.44</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Williams et al</td>
<td>-0.37</td>
<td>-1.23 to 0.48</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Williams et al</td>
<td>0.42</td>
<td>-0.46 to 1.29</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Williams et al</td>
<td>0.78</td>
<td>-0.12 to 1.67</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Wims et al</td>
<td>0.20</td>
<td>-0.34 to 0.74</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>POOLED</td>
<td>0.81</td>
<td>0.59-1.04</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4** Effects of cognitive behavior therapy (CBT) for panic disorder compared to control conditions: forest plot.

As shown in Table 2, very few predictors were significant in these analyses, possibly because of the relatively small number of studies per disorder and the relatively large number of predictors.

**DISCUSSION**

In this study, we aimed to establish the most up-to-date and accurate estimate of the effects of CBT in the treatment of MDD, GAD, PAD and SAD. We also aimed to examine whether the problems of publication bias, low quality of trials, and the use of waiting list control groups have an impact on the effect
sizes. We found that the overall effects for all four disorders were large, ranging from g = 0.75 for MDD to g = 0.80 for GAD, g = 0.81 for PAD, and g = 0.88 for SAD.

The first problem, publication bias, mostly affected the outcomes of CBT for GAD and MDD. For GAD, it was estimated that about one quarter of the studies were missing and, after adjusting for these missing studies, the effect size dropped from g = 0.80 to g = 0.59. For MDD, 14% of the studies were missing, and the pooled effect size dropped from g = 0.75 to g = 0.65. However, this was a relatively small drop compared to that reported in other studies on publication bias in psychotherapies for MDD15,18,184. This may be due to the fact that we used more stringent inclusion criteria for this meta-analysis (only patients meeting diagnostic criteria for MDD; only wait-listing list, treatment-as-usual or pill placebo control groups; only CBT). In PAD and SAD, we found few indications of publication bias.

The second problem we aimed to examine was the quality of trials. We found that the methodological quality in most studies was low or unknown. We evaluated the quality by the Cochrane “risk of bias” assessment tool, and found that across all disorders only 25 trials (17.4%) were rated as high-quality. The effect size was lower in high-quality studies for PAD (g = 0.61 compared to g = 0.81 in all studies) and SAD (g = 0.76 compared to g = 0.88 in all studies). We did not find strong indications that the quality of trials was associated with the effect size in MDD and GAD. Although we did not find a strong association between effect size and quality of trials for all disorders, the small number of high-quality studies still means that the overall effect sizes we found for all four disorders are uncertain.

The third problem we aimed to examine was the influence of waiting list control groups on the effects of CBT. We found that the vast majority of studies for the three anxiety disorders used a waiting list control group (77.4% of the comparisons for GAD, 78.6% for PAD, and 83.3% for SAD). In MDD, the number of studies using care-as-usual and pill placebo control conditions was larger, but still 44.4% (28 out of 63) of the included studies used a waiting list control group. This means that much of the evidence on the effects of CBT is based on the use of waiting list control groups. As indicated earlier, improvements found in patients on waiting lists are lower than can be expected on the basis of spontaneous remission19,185. Waiting list is probably a “nocebo”21, considerably overestimating the effects of psychological treatments. This was confirmed in our meta-analysis, in which we found for each of the disorders that studies with a waiting list control group resulted in significantly higher effect sizes than those with a care-as-usual or pill placebo control group.

The few studies on anxiety disorders that used care-as-usual or pill placebo control groups indicated small to moderate

<p>| Table 2 Standardized regression coefficients of characteristics of studies on cognitive behavior therapy for major depression (MDD), generalized anxiety disorder (GAD), panic disorder (PAD) and social anxiety disorder (SAD) compared to control conditions |
|-----------------------------------------------|----------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>MDD</th>
<th>GAD</th>
<th>PAD</th>
<th>SAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of trial</td>
<td>−0.05</td>
<td>0.07</td>
<td>0.46</td>
<td>−0.01</td>
</tr>
<tr>
<td>Control condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting list</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care-as-usual</td>
<td>−0.43</td>
<td>0.15</td>
<td>0.01</td>
<td>−0.30</td>
</tr>
<tr>
<td>Pill placebo</td>
<td>−0.44</td>
<td>0.30</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>Adults vs. specific target groups</td>
<td>0.01</td>
<td>0.17</td>
<td>0.95</td>
<td>−0.45</td>
</tr>
<tr>
<td>Format</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>−0.23</td>
<td>0.21</td>
<td>0.28</td>
<td>−0.17</td>
</tr>
<tr>
<td>Guided self-help</td>
<td>−0.32</td>
<td>0.23</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>Mixed/other</td>
<td>−0.28</td>
<td>0.28</td>
<td>0.32</td>
<td>0.04</td>
</tr>
<tr>
<td>Number of sessions</td>
<td>−0.01</td>
<td>0.02</td>
<td>0.67</td>
<td>−0.01</td>
</tr>
<tr>
<td>Geographic area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>−0.02</td>
<td>0.19</td>
<td>0.92</td>
<td>−0.51</td>
</tr>
<tr>
<td>Australia</td>
<td>0.31</td>
<td>0.29</td>
<td>0.29</td>
<td>−0.19</td>
</tr>
<tr>
<td>Other</td>
<td>0.47</td>
<td>0.22</td>
<td>0.04</td>
<td>−0.78</td>
</tr>
</tbody>
</table>

Significant p values are highlighted in bold prints.
effect sizes. In the four studies comparing CBT for PAD with care-as-usual, the effect size was even non-significant (p=0.17). Furthermore, because of the small number of studies, and the even smaller number of high-quality studies, the effects of CBT in anxiety disorders are quite uncertain.

An exception to the small to moderate effects of CBT in anxiety disorders was the group of studies comparing CBT to pill placebo for GAD. These studies resulted in a very large effect size (g=1.32). However, because of the small number of trials and the low quality of all three of them, these results should be considered with caution.

One reason to conduct this meta-analysis was to examine whether the quality of trials has increased in recent years. Indeed, 29.5% of the studies conducted in 2010 or later were rated as high-quality, while that was true for only 12.0% of the older studies. Furthermore, 52.0% of all high-quality studies were conducted in 2010 or later. This is likely to have led to a more accurate estimate of effect sizes.

The present study has several strengths, including the broad scope of the meta-analyses, covering four common mental disorders, the rigorous selection and assessment of the trials, and their relatively large number.

One possible limitation is that we used strict inclusion criteria, only focusing on trials in which patients met diagnostic criteria for the disorder according to a structured interview and trials in which either a waiting list, care-as-usual or pill placebo control group was used. We did not include studies in which, for example, generic counselling was used as a control condition. This may contribute to explain the small number of trials comparing CBT with control conditions other than waiting lists, especially in anxiety disorders and among the sets of high-quality studies. Furthermore, care-as-usual control groups can vary considerably depending on the country and the treatment setting where the therapy is offered, and may therefore be too heterogeneous to allow a reliable assessment of the effects across studies. Finally, we only focused on short-term outcomes, because only few studies reported long-term outcomes and the follow-up periods differed considerably.

On the basis of our data, we conclude that CBT is probably effective in the treatment of MDD, GAD, PAD and SAD, and that the effects are large when compared to waiting list control groups, but small to moderate when compared to more conservative control groups, such as care-as-usual and pill placebo. Because of the small number of high-quality studies, these effects are still uncertain and should be considered with caution.

REFERENCES


91. Wong DFK. Cognitive behavioral treatment groups for people with chronic depression in Hong Kong: a randomized wait-list control design. Depress Anxiety 2008;25:142-8.


185. Cuijpers P, Cristea IA. What if a placebo effect explained all the activity of depression treatments? World Psychiatry 2015;14:310-1

DOI:10.1002/wps.20346